

Substitute Sequence Listing_USSN 10587032_PP019482.007
SUBSTITUTE SEQUENCE LISTING

<110> CHIRON CORPORATION
HARDY, Stephen F
DONNELLY, III, John J
ZUR MEGEDE, Jan T

<120> VECTORS FOR EXPRESSION OF HML-2 POLYPEPTIDES

<130> PP19482.0007

<140> 10/587,032
<141> 2006-07-24

<150> PCT/US03/18666
<151> 2003-06-13

<150> 60/388831
<151> 2002-06-16

<150> 60/472189
<151> 2003-05-20

<160> 83

<170> PatentIn, version 3.5

<210> 1

<211> 1998

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 1
atggggcaaa ctaaaagtaa aattaaaagt aaatatgcct ctatctcag ctttattaaa 60
attcttttaa aaagagggggg agttaaaagta tctacaaaaa atctaataa gctattcaa 120
ataatagaac aattttgccc atggttcca gaacaaggaa ctttagatct aaaagattgg 180
aaaagaattg taaaggaaact aaaacaagca ggttagaagg gtaatcat tccacttaca 240
gtatggaatg attggggccat tattaaagca gctttagaac catttcaaac agaagaagat 300
agcgtttcag tttctgtatgc cccttgaagc tgtataatag attgtatga aaacacaagg 360
aaaaaatccc agaaagaaac ggaaggtttta cattgcgaat atgttagcaga gccggtaatg 420
gctcagtcaa cgcaaaatgt tgactataat caattacagg aggtatata tcctgaaacg 480
ttaaaattag aaggaaaagg tccagaatta gtggggccat cagatctaa accacgaggc 540
acaagtccct ttccagcagg tcaggtgcct gtaacattac aacctcaaaa gcaggttaaa 600
gaaaataaga cccaaaccgcc agtagcctat caataactggc ctccggctga acttcagtat 660
cggccacccc cagaaagtca gtatggatat ccaggaatgc ccccgacacc acagggcagg 720
gcccataacc ctcagccgcc cactaggaga cttaatccctt cggcaccacc tagtagacag 780
ggtagtaaat tacatgaaat tattgataaa tcaagaaaagg aaggagatac tgaggcatgg 840
caattcccg taacgttaga accgatgcca cctggagaag gagcccaaga gggagagcct 900
cccacagttg aggccagata caagtcttt tcgataaaaa agctaaaaga tatgaaagag 960
ggagtaaaac agtatggacc caactccct tatatgagga cattattaga ttccattgct 1020
catggacata gactcattcc ttatgattgg gagattctgg caaaatcgtc tctctcaccc 1080
tctcaatttt tacaattttaa gacttggtgg attgatgggg tacaagaaca ggtccgaaga 1140
aatagggctg ccaatccctcc agttaacata gatgcagatc aactattagg aataggtcaa 1200
aattggatgtt ctattagtca acaagcattt atgcaaaatgt aggcatttgca gcaagttaga 1260
gctatctgcc tttagagcctg ggaaaaaaatc caagaccagg gaagtacctg cccctcattt 1320
aatacagtaa gacaagggttcc aaaagagccc tattcctgatt ttgtggcaag gctccaagat 1380
gttgctcaaa agtcaatttgc tgatgaaaaa gcccgttaagg tcatatgtgg gttgatggca 1440
tatgaaaacg ccaatccctga gtgtcaatca gccattaagc cattaaaagg aaaggttcct 1500
gcaggatcag atgttatctc agaatatgtt aaggttgcct atgaaatcggtt aggagctatg 1560
ctaaagcta tgcttatggc tcaagcaata acaggatgtt ttttaggagg acaagttaga 1620
acatttgaa gaaaatgttta taattgttgtt caaattggtc acttaaaaaaa gaattgcccc 1680
gtcttaataa aacagaatata aactattcaa gcaactacaa caggtagaga gccacctgac 1740

Substitute Sequence Listing_USSN 10587032_PP019482.007

ttatgtccaa	gatgtaaaaa	aggaaaacat	tgggctagtc	aatgtcggtc	taaatttgat	1800
aaaaatgggc	aaccattgtc	gggaaacgag	caaaggggcc	agcctcaggc	cccacaacaa	1860
actggggcat	tcccaattca	gccatttgtt	cctcagggtt	ttcagggaca	acaaccccca	1920
ctgtcccaag	tgtttcaggg	aataagccag	ttaccacaat	acaacaattg	tccccgcca	1980
caagcggcag	tgcagcag					1998

<210> 2

<211> 2001

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 2

atggggcaaa	ctaaaagtaa	aattaaaagt	aaatatgcct	cttatctcag	ctttattaaa	60
attcttttaa	aaagaggggg	agttaaaagta	tctacaaaaa	atctaatacaa	gctatttcaa	120
ataatagaac	aattttgcc	atggggttcca	gaacaaggaa	ctttagatct	aaaagattgg	180
aaaagaattg	gtaaaggaaact	aaaacaagca	ggtaggaagg	gtaatatcat	tccacttaca	240
gtatggatg	attggggccat	tattaaagca	gctttagaac	catttcaaac	agaagaagat	300
agcgtttag	tttctgtatgc	cccttggaaagc	tgtataatag	attgtaatga	aaacacaagg	360
aaaaaatccc	agaaaagaaac	ggaagggtta	catttgcgaat	atgttagcaga	gccggtaatg	420
gctcagtcaa	cgcaaaatgt	tgactataat	caatttacagg	agggtatata	tcctgaaacg	480
ttaaaattag	aaggaaaagg	tccagaattt	gtggggccat	cagatctaa	accacgaggc	540
acaagtccct	ttccagcagg	tcaggtgcct	gtaacattac	aacctcaaaa	gcaggttaaa	600
gaaaataaga	cccaaccgccc	agttagcctat	caataacttgc	ctccggctga	acttcagtat	660
cggccacccc	cagaaagtca	gtatggatat	ccaggaatgc	ccccagcacc	acagggcagg	720
gcccatacc	ctcagccgccc	cactaggaga	cttaatccta	cgccaccacc	tagtagacag	780
ggtagtaaat	tacatgaaat	tattgataaa	tcaagaagg	aaggagatac	tgaggcatgg	840
caattcccg	taacgttaga	accgtatgca	ccttggaaag	gagccccaga	gggagagcct	900
cccacagtt	aggccagata	caagtctttt	tcgataaaaa	agctgaaaga	tatgaaagag	960
ggagtagaaac	agtatggacc	caactccct	tatatgagga	cattattaga	ttccattgct	1020
catggacata	gactcattcc	ttatgattgg	gagattctgg	caaaatcgtc	tctctcaccc	1080
tctcaatttt	tacaatttaa	gacttggtgg	attgtatggg	tacaagaaca	ggtccgaaga	1140
aatagggctg	ccaatccccc	agttaaacata	gatgcagatc	aactattagg	aataggtcaa	1200
aattggagta	ctttagtca	acaagcatta	atgcaaaatg	aggccattga	gcaagttaga	1260
gctatctgcc	ttagagcctg	ggaaaaaaatc	caagaccagg	gaagtacctg	ccccctcattt	1320
aatacagtaa	gacaagggtc	aaaagagccc	tatcctgatt	ttgtggcaag	gctccaagat	1380
gttgctcaaa	agtcaattgc	tgatgaaaaa	gccccgttaagg	tcatagtgg	gtttagtggca	1440
tatgaaaacg	ccaaatccctga	gtgtcaatca	gccatttaagc	cattaaaagg	aaaggttcct	1500
gcaggatcag	atgtatctc	agaatatgt	aaaggctgt	atgaaatcgg	aggagctatg	1560
tataaagcta	tgctttaggc	tcaagcaata	acaggagttt	ttttaggagg	acaagttaga	1620
acatttggaa	gaaaatgtt	taatttgtgt	caaattgttc	actaaaaaaa	gaattgccc	1680
gtcttaaata	aacagaatata	aacttattca	gcaactcaa	cagtagaga	gccacctgac	1740
ttatgtccaa	gatgtaaaaa	aggaaaacat	tgggctagtc	aatgtcggtc	taaatttgat	1800
aaaaatgggc	aaccattgtc	gggaaacgag	caaaggggcc	agcctcaggc	cccacaacaa	1860
actggggcat	tcccaattca	gccatttgtt	cctcagggtt	ttcagggaca	acaaccccca	1920
ctgtcccaag	tgtttcaggg	aataagccag	ttaccacaat	acaacaattg	tccccgcca	1980
caagcggcag	tgcagcagta	g				2001

<210> 3

<211> 2004

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 3

atggggcaaa	ctaaaagtaa	aactaaaagt	aaatatgcct	cttatctcag	ctttattaaa	60
attcttttaa	aaagaggggg	agtttagagta	tctacaaaaa	atctaatacaa	gctatttcaa	120
ataatagaac	aattttgcc	atggggttcca	gaacaaggaa	ctttagatct	aaaagattgg	180
aaaagaattg	gcgaggaaact	aaaacaagca	ggtagaaagg	gtaatatcat	tccacttaca	240
gtatggatg	attggggccat	tattaaagca	gctttagaac	catttcaaac	aaaagaagat	300
agcgtttag	tttctgtatgc	cccttggaaagc	tgtgtatag	attgtaatga	aaagacagg	360
agaaaatccc	agaaaagaaac	agaaagtta	catttgcgaat	atgttaacaga	gccagttatg	420
gctcagtcaa	cgcaaaatgt	tgactataat	caatttacagg	gggtatata	tcctgaaacg	480
ttaaaattag	aaggaaaagg	tccagaattt	gtggggccat	cagatctaa	accacgagg	540
ccaagtccct	ttccagcagg	tcaggtgcct	gtaacattac	aacctcaaac	gcaggttaaa	600

Substitute Sequence Listing_USSN 10587032_PP019482.007

gaaaataaga	cccaaccgcc	agtagcttat	caatactggc	cgcggctga	acttcagttat	660
ctgccacccc	cagaaagtca	gtatggatat	ccaggaatgc	ccccagcact	acagggcagg	720
gcgcacatac	ctcagccgcc	caactgtgaga	cttaatccta	cagcatcactg	tagtggacaa	780
gggtgtacac	tgcacgcagt	cattgtgaa	gccagaaaac	aggagatct	tgaggcatgg	840
cggttccctgg	taattttaca	actgttacag	gccggggaaag	agactcaagt	aggagcgct	900
gccc gagctg	agactagatg	tgaaccttgc	accatgaaaa	tgttaaaaaga	tataaaggaa	960
ggagttaaac	aatatggatc	caactcccct	tatataagaa	cattattaga	ttccattgct	1020
catggaaata	gacttactcc	ttatgactgg	gaaagtttg	ccaaatcttc	cctttcatcc	1080
tctcagttatc	tacagtttaa	aacctggtgg	attgtatggag	tacaagaaca	ggtacgaaaa	1140
aatcaggcta	ctaagcccac	tgttaatata	gacgcagacc	aattgtttagg	aacaggtcca	1200
aattggagca	ccattaacca	acaatcagtg	atgcagaatg	aggctattga	acaagtaagg	1260
gctatttgcc	tcagggccctg	gggaaaaatt	caggaccctg	gaacagctt	cccttataat	1320
tcaatttagac	aaggctctaa	agagccatat	cctgactttg	tggcaagatt	acaagatgct	1380
gctcaaagt	ctattacaga	tgacaatgcc	cggaaaagtt	ttttagaatt	aatggcttat	1440
gaaaatgcaa	atccagaatg	tcagtcggcc	ataaagccat	taaaaggaaa	agttccagca	1500
ggagttgatg	taattacaga	atatgtgaa	gcttgtgatg	ggatttgagg	agctatgcat	1560
aaggcaatgc	taatggctca	agcaatgagg	gggctcactc	taggaggaca	agttagaaca	1620
tttggaaaaa	aatgttataa	tttgttcaaa	atcggtcatc	tgaaaaggag	ttgcccagtc	1680
ttaaataaac	agaatataat	aaatcaagct	attacagcaa	aaaataaaaaa	gccatctggc	1740
ctgtgtccaa	aatgtggaaa	aggaaaacat	tggggccaaatc	aatgtcattc	taaatttgat	1800
aaagatgggc	aaccattgtc	gggaaacagg	aagaggggccc	agcctcaggc	cccccaacaa	1860
actggggcat	tcccagttca	actgttgg	cctcagggtt	ttcaaggaca	acaaccctta	1920
cagaaaatac	caccacttca	gggagtcagc	caattacaac	aatccaacag	ctgtcccgcg	1980
ccacagcagg	cagcggcaca	gtag				2004

<210> 4

<211> 852

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 4

atggggcaaa	ctaaaagtaa	aattaaaagt	aaatatgcct	cttatctcag	ctttattaaa	60
attcttttaa	aaagaggggg	agttaaagta	tctacaaaaa	atctaataa	gctatttcaa	120
ataatagaac	aattttgccc	atggtttcca	gaacaagaa	cttcagatct	aaaagattgg	180
aaaagaattg	gtaaggaact	aaaacaagca	ggtaggaagg	gtaatatcat	tccacttaca	240
gtatgaaatg	attggggccat	tattaaagca	gctttagaac	catttcaaac	agaagaagat	300
agcatttcag	tttctgtatgc	cccttggaa	tgtttaatag	attgtatga	aaacacaagg	360
aaaaaatccc	agaaagaaac	cgaagttt	catttgcgaat	atgttagcaga	gccggtaatg	420
gctcagtc	cgcaaaaatgt	tgactataat	caatttacagg	aggttatata	tcctgaaacg	480
ttaaaattag	aggaaaaagg	tccagaatta	atggggccat	cagagtctaa	accacgaggc	540
acaagtccct	ttccagcagg	tcagtgctc	gtaagattac	aacctaaaaa	gcaggttaaa	600
gaaaataaga	cccaaccgca	agtagcctat	caatactgcc	gctggctgaa	cttcagttatc	660
ggccacccccc	agaaagtcag	tatggatatc	caggaatgcc	cccaagcacca	cagggcagg	720
cgcctatcca	tcagccgccc	actaggagac	ttaatcctat	ggcaccacct	agtagacagg	780
gtatgtatt	acatgaaatt	attgataat	caagaaagga	aggagatact	gaggcatggc	840
aattcccagt	aa					852

<210> 5

<211> 666

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 5

Met	Gly	Gln	Thr	Lys	Ser	Lys	Ile	Lys	Ser	Lys	Tyr	Ala	Ser	Tyr	Leu
1															

Ser	Phe	Ile	Lys	Ile	Leu	Leu	Lys	Arg	Gly	Gly	Val	Lys	Val	Ser	Thr

Lys	Asn	Leu	Ile	Lys	Leu	Phe	Gln	Ile	Ile	Glu	Gln	Phe	Cys	Pro	Trp

Phe	Pro	Glu	Gln	Gly	Thr	Leu	Asp	Leu	Lys	Asp	Trp	Lys	Arg	Ile	Gly

50

Substitute Sequence Listing_USSN 10587032_PP019482.007
55 60

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65 70 75 80

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
85 90 95

Thr Glu Glu Asp Ser Val Ser Val Asp Ala Pro Gly Ser Cys Ile
100 105 110

Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
115 120 125

Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
130 135 140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145 150 155 160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
165 170 175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
180 185 190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
195 200 205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
210 215 220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225 230 235 240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
245 250 255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
260 265 270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
275 280 285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
290 295 300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305 310 315 320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
325 330 335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
340 345 350

Leu Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
355 360 365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
370 375 380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln

385 Substitute Sequence Listing_USSN 10587032_PP019482.007
 390 395 400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
405 410 415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
420 425 430

Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
435 440 445

Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
450 455 460

Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
465 470 475 480

Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
485 490 495

Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
500 505 510

Cys Asp Gly Ile Gly Gly Ala Met Tyr Lys Ala Met Leu Met Ala Gln
515 520 525

Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg
530 535 540

Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro
545 550 555 560

Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg
565 570 575

Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala
580 585 590

Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly
595 600 605

Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
610 615 620

Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro
625 630 635 640

Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn
645 650 655

Cys Pro Pro Pro Gln Ala Ala Val Gln Gln
660 665

<210> 6

<211> 667

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 6

Met Gly Gln Thr Lys Ser Lys Thr Lys Ser Lys Tyr Ala Ser Tyr Leu
1 5 10 15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Arg Val Ser Thr
Page 5

Substitute Sequence Listing_USSN 10587032_PP019482.007
20 25 30

Lys Asn Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
 35 40 45
 Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
 50 55 60
 Glu Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
 65 70 75 80
 Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
 85 90 95
 Thr Lys Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Val
 100 105 110
 Ile Asp Cys Asn Glu Lys Thr Gly Arg Lys Ser Gln Lys Glu Thr Glu
 115 120 125
 Ser Leu His Cys Glu Tyr Val Thr Glu Pro Val Met Ala Gln Ser Thr
 130 135 140
 Gln Asn Val Asp Tyr Asn Gln Leu Gln Gly Val Ile Tyr Pro Glu Thr
 145 150 155 160
 Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
 165 170 175
 Lys Pro Arg Gly Pro Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
 180 185 190
 Leu Gln Pro Gln Thr Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
 195 200 205
 Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Leu Pro Pro Pro
 210 215 220
 Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Leu Gln Gly Arg
 225 230 235 240
 Ala Pro Tyr Pro Gln Pro Pro Thr Val Arg Leu Asn Pro Thr Ala Ser
 245 250 255
 Arg Ser Gly Gln Gly Gly Thr Leu His Ala Val Ile Asp Glu Ala Arg
 260 265 270
 Lys Gln Gly Asp Leu Glu Ala Trp Arg Phe Leu Val Ile Leu Gln Leu
 275 280 285
 Val Gln Ala Gly Glu Glu Thr Gln Val Gly Ala Pro Ala Arg Ala Glu
 290 295 300
 Thr Arg Cys Glu Pro Phe Thr Met Lys Met Leu Lys Asp Ile Lys Glu
 305 310 315 320
 Gly Val Lys Gln Tyr Gly Ser Asn Ser Pro Tyr Ile Arg Thr Leu Leu
 325 330 335
 Asp Ser Ile Ala His Gly Asn Arg Leu Thr Pro Tyr Asp Trp Glu Ser
 340 345 350
 Leu Ala Lys Ser Ser Leu Ser Ser Ser Gln Tyr Leu Gln Phe Lys Thr

Substitute Sequence Listing_USSN 10587032_PP019482.007
355 360 365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Lys Asn Gln Ala Thr
370 375 380

Lys Pro Thr Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Thr Gly Pro
385 390 395 400

Asn Trp Ser Thr Ile Asn Gln Gln Ser Val Met Gln Asn Glu Ala Ile
405 410 415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Gly Lys Ile Gln Asp
420 425 430

Pro Gly Thr Ala Phe Pro Ile Asn Ser Ile Arg Gln Gly Ser Lys Glu
435 440 445

Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Ala Ala Gln Lys Ser
450 455 460

Ile Thr Asp Asp Asn Ala Arg Lys Val Ile Val Glu Leu Met Ala Tyr
465 470 475 480

Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys Gly
485 490 495

Lys Val Pro Ala Gly Val Asp Val Ile Thr Glu Tyr Val Lys Ala Cys
500 505 510

Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln Ala
515 520 525

Met Arg Gly Leu Thr Leu Gly Gln Val Arg Thr Phe Gly Lys Lys
530 535 540

Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Arg Ser Cys Pro Val
545 550 555 560

Leu Asn Lys Gln Asn Ile Ile Asn Gln Ala Ile Thr Ala Lys Asn Lys
565 570 575

Lys Pro Ser Gly Leu Cys Pro Lys Cys Gly Lys Gly Lys His Trp Ala
580 585 590

Asn Gln Cys His Ser Lys Phe Asp Lys Asp Gly Gln Pro Leu Ser Gly
595 600 605

Asn Arg Lys Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
610 615 620

Pro Val Gln Leu Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Leu
625 630 635 640

Gln Lys Ile Pro Pro Leu Gln Gly Val Ser Gln Leu Gln Gln Ser Asn
645 650 655

Ser Cys Pro Ala Pro Gln Gln Ala Ala Pro Gln
660 665

<210> 7
<211> 283

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)
Page 7

Substitute Sequence Listing_USSN 10587032_PP019482.007

<400> 7
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1 5 10 15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
20 25 30

Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
35 40 45

Phe Pro Glu Gln Gly Thr Ser Asp Leu Lys Asp Trp Lys Arg Ile Gly
50 55 60

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65 70 75 80

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
85 90 95

Thr Glu Glu Asp Ser Ile Ser Val Ser Asp Ala Pro Gly Ser Cys Leu
100 105 110

Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
115 120 125

Ser Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
130 135 140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145 150 155 160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Met Gly Pro Ser Glu Ser
165 170 175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Leu Val Arg
180 185 190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Gln Val
195 200 205

Ala Tyr Gln Tyr Cys Arg Trp Leu Asn Phe Ser Ile Gly His Pro Gln
210 215 220

Lys Val Ser Met Asp Ile Gln Glu Cys Pro Gln His His Arg Ala Gly
225 230 235 240

Arg His Thr Ile Ser Arg Pro Leu Gly Asp Leu Ile Leu Trp His His
245 250 255

Leu Val Asp Arg Val Val Asn Tyr Met Lys Leu Leu Ile Asn Gln Glu
260 265 270

Arg Lys Glu Ile Leu Arg His Gly Asn Ser Gln
275 280

<210> 8

<211> 434

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 8

Met Pro Pro Ala Pro Gln Gly Arg Ala Pro Tyr His Gln Pro Pro Thr

1

Substitute Sequence Listing_USSN 10587032_PP019482.007
5 10 15

Arg Arg Leu Asn Pro Met Ala Pro Pro Ser Arg Gln Gly Ser Glu Leu
20 25 30

His Glu Ile Ile Asp Lys Ser Arg Lys Glu Gly Asp Thr Glu Ala Trp
35 40 45

Gln Phe Pro Val Thr Leu Glu Pro Met Pro Pro Gly Glu Gly Ala Gln
50 55 60

Glu Gly Glu Pro Pro Thr Val Glu Ala Arg Tyr Lys Ser Phe Ser Ile
65 70 75 80

Lys Met Leu Lys Asp Met Lys Glu Gly Val Lys Gln Tyr Gly Pro Asn
85 90 95

Ser Pro Tyr Met Arg Thr Leu Leu Asp Ser Ile Ala Tyr Gly His Arg
100 105 110

Leu Ile Pro Tyr Asp Trp Glu Ile Leu Ala Lys Ser Ser Leu Ser Pro
115 120 125

Ser Gln Phe Leu Gln Phe Lys Thr Trp Trp Ile Asp Gly Val Gln Glu
130 135 140

Gln Val Arg Arg Asn Arg Ala Ala Asn Pro Pro Val Asn Ile Asp Ala
145 150 155 160

Asp Gln Leu Leu Gly Ile Gly Gln Asn Trp Ser Thr Ile Ser Gln Gln
165 170 175

Ala Leu Met Gln Asn Glu Ala Ile Glu Gln Val Arg Ala Ile Cys Leu
180 185 190

Arg Ala Trp Glu Lys Ile Gln Asp Pro Gly Ser Thr Cys Pro Ser Phe
195 200 205

Asn Thr Val Arg Gln Gly Ser Lys Glu Pro Tyr Pro Asp Phe Val Ala
210 215 220

Arg Leu Gln Asp Val Ala Gln Lys Ser Ile Ala Asp Glu Lys Ala Gly
225 230 235 240

Lys Val Ile Val Glu Leu Met Ala Tyr Glu Asn Ala Asn Pro Glu Cys
245 250 255

Gln Ser Ala Ile Lys Pro Leu Lys Gly Lys Val Pro Ala Gly Ser Asp
260 265 270

Val Ile Ser Glu Tyr Val Lys Ala Cys Asp Gly Ile Gly Gly Ala Met
275 280 285

His Lys Ala Met Leu Met Ala Gln Ala Ile Thr Gly Val Val Leu Gly
290 295 300

Gly Gln Val Arg Thr Phe Gly Gly Lys Cys Tyr Asn Cys Gly Gln Ile
305 310 315 320

Gly His Leu Lys Lys Asn Cys Pro Val Leu Asn Lys Gln Asn Ile Thr
325 330 335

Ile Gln Ala Thr Thr Gly Arg Glu Pro Pro Asp Leu Cys Pro Arg

Substitute Sequence Listing_USSN 10587032_PP019482.007
340 345 350

Cys Lys Lys Gly Lys His Trp Ala Ser Gln Cys Arg Ser Lys Phe Asp
355 360 365
Lys Asn Gly Gln Pro Leu Ser Gly Asn Glu Gln Arg Gly Gln Pro Gln
370 375 380
Ala Pro Gln Gln Thr Gly Ala Phe Pro Ile Gln Pro Phe Val Pro Gln
385 390 395 400
Gly Phe Gln Gly Gln Gln Pro Pro Leu Ser Gln Val Phe Gln Gly Ile
405 410 415
Ser Gln Leu Pro Gln Tyr Asn Asn Cys Pro Ser Pro Gln Ala Ala Val
420 425 430
Gln Gln

<210> 9
<211> 666
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 9
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1 5 10 15
Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
20 25 30
Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
35 40 45
Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
50 55 60
Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65 70 75 80
Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
85 90 95
Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
100 105 110
Ile Asp Cys Asn Glu Asn Thr Gly Lys Lys Ser Gln Lys Glu Thr Glu
115 120 125
Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
130 135 140
Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145 150 155 160
Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
165 170 175
Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
180 185 190
Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val

195 Substitute Sequence Listing_USSN 10587032_PP019482.007
200 205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
210 215 220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225 230 235 240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
245 250 255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
260 265 270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
275 280 285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
290 295 300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305 310 315 320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
325 330 335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
340 345 350

Gln Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
355 360 365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
370 375 380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln
385 390 395 400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
405 410 415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
420 425 430

Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
435 440 445

Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
450 455 460

Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala
465 470 475 480

Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys
485 490 495

Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala
500 505 510

Cys Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln
515 520 525

Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg
Page 11

Substitute Sequence Listing_USSN 10587032_PP019482.007
 530 535 540

Lys	Cys	Tyr	Asn	Cys	Gly	Gln	Ile	Gly	His	Leu	Lys	Lys	Asn	Cys	Pro	
545					550					555					560	
Val	Leu	Asn	Lys	Gln	Asn	Ile	Thr	Ile	Gln	Ala	Thr	Thr	Thr	Gly	Arg	
						565			570					575		
Glu	Pro	Pro	Asp	Leu	Cys	Pro	Arg	Cys	Lys	Lys	Gly	Lys	His	Trp	Ala	
						580		585						590		
Ser	Gln	Cys	Arg	Ser	Lys	Phe	Asp	Lys	Asn	Gly	Gln	Pro	Leu	Ser	Gly	
					595		600				605					
Asn	Glu	Gln	Arg	Gly	Gln	Pro	Gln	Ala	Pro	Gln	Gln	Thr	Gly	Ala	Phe	
					610		615				620					
Pro	Ile	Gln	Pro	Phe	Val	Pro	Gln	Gly	Phe	Gln	Gly	Gln	Gln	Pro	Pro	
					625		630			635				640		
Leu	Ser	Gln	Val	Phe	Gln	Gly	Ile	Ser	Gln	Leu	Pro	Gln	Tyr	Asn	Asn	
					645			650					655			
Cys	Pro	Pro	Pro	Gln	Ala	Ala	Val	Gln	Gln							
					660			665								

<210> 10
 <211> 1000

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 10

atggccaacc	attgtcggga	aacgagcaaa	ggggccagcc	tcaggccccca	caacaactg	60
gggcattccc	aattcagcca	tttggcctc	agggtttca	gggacaacaa	ccccactgt	120
cccaagtgtt	tcagggata	agccagttac	cacaatacaa	caattgtccc	ccgcccacaag	180
cggcagtgc	gcagtagatt	tatgtactat	acaaggcgtc	tctctgcttc	caggggagcc	240
cccacaaaaa	acccccacag	gggtatatgg	accctgcct	aaggggactg	taggactaat	300
cttggacga	tcaagtctaa	atctaaaagg	agttcaatt	catactagt	tggttgattc	360
agactataaa	ggcggaaattc	aattggttat	tagctctca	attcttggaa	gtgccagtcc	420
aagagacagg	attgctcaat	tattactcct	gccatacatt	aagggtggaa	atagtgaat	480
aaaaagaata	ggagggcttg	gaagcactga	tccaaacagga	aaggctgcat	attgggcaag	540
tcaggtctca	gagaacagac	ctgtgtgtaa	ggccattatt	caaggaaaac	agtttgaagg	600
gttggtagac	actggagcag	atgtctctat	cattgcttta	aattcagtgcc	caaaaaattt	660
gcctaaacaa	aaggctgtt	caggacttgt	cgccataggc	acagcctcag	aagtgtatca	720
aagtacggag	attttacatt	gcttagggcc	agataatcaa	gaaagtactg	ttcagccaat	780
gattacttca	attcctctta	atctgtgggg	tcgagattna	ttacaacaat	ggggtgccga	840
aatcaccatg	cccgctccat	cata>tagccc	cacgagtc	aaaatcatga	ccaagatggg	900
atatatacc	ggaaagggac	tagggaaaaa	tgaagatggc	attaaaattc	cagttgaggc	960
taaaataaaat	caagaaagag	aaggaatagg	gaatccttgc			1000

<210> 11

<211> 1004

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 11

atggccaacc	attgtcggga	aacaggaaga	ggggccagcc	tcaggccccca	caacaactg	60
gggcattccc	agttcaactg	tttggcctc	agggtttca	aggacaacaa	ccccctacaga	120
aaataccacc	acttcaggg	gtcagccat	tacaacaatc	caacagctgt	cccgcccac	180
agcaggcagc	gccacagttag	atttatgttc	cacccaaatg	gtctcttac	tccctggaga	240
gccccccacaa	aaagattcc	gaggggtata	tggcccgtg	ccagaaggga	gggttaggc	300
tatttiaggg	agatcaagtc	taaatttga	gggagtc	attcatactg	gggttaattt	360
ttcagattat	aaagggggaa	ttcagttgt	gatcagctcc	actgtccct	ggagtgc	420

Substitute Sequence Listing_USSN 10587032_PP019482.007

tccagggtat	agaattgctc	aattactgct	tttgcccttat	gttaaaaattt	ggaaaaacaa	480
aacggaaaga	acaggaggtt	ttggaagtac	caaccctgca	ggaaaagcca	cttattggc	540
taatcaggta	tcagaggata	gaccctgtg	tacagtca	attcaggaa	agagttgaa	600
ggatttagtgg	ataccaggc	tgatgttct	atcatcgca	taggcaccgc	ctcagaagtg	660
tatcaaagt	ccatgattt	acattgtcta	ggatctgata	atcaagaaag	tacggttcag	720
cctatgatca	cttctattcc	aatcaattt	tggggccag	acttgtaca	acaatggcat	780
gcagagatta	ctatcccagc	ctccctatac	agccccagga	atcaaaaaat	catgactaaa	840
atgggatagc	tccctaaaaa	gggacttagg	aagaatgaag	atggcattaa	agtcccaact	900
gaggctgaaa	aaaatcaaaa	aaagaaaaagg	aatagggcat	cctttttaga	agcggtcact	960
gtagagcctc	caaaacccat	tccattaatt	tggggggaaa	aaaa		1004

<210> 12
<211> 279
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 12
atggagattt tacattgctt agggccagat aatcaagaaa gtactgttca gccaatgatt 60
acttcaattc ctcttaatct gtggggtcga gatttattac aacaatgggg tgcggaaatc 120
accatgcccgt cttcattata tagccccacg agtcaaaaaa tcatgaccaa gatggatata 180
ataccaggaa agggacttagg gaaaaatgaa gatggcatta aagttccagt tgaggctaaa 240
ataaatcaag aaagagaagg aatagggtat ccttttag 279

<210> 13
<211> 92
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 13
Met Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser Thr Val
1 5 10 15
Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg Asp Leu
20 25 30
Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Leu Tyr Ser
35 40 45
Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro Gly Lys
50 55 60
Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Val Pro Val Glu Ala Lys
65 70 75 80
Ile Asn Gln Glu Arg Glu Gly Ile Gly Tyr Pro Phe
85 90

<210> 14
<211> 333
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 14
Trp Ala Thr Ile Val Gly Lys Arg Ala Lys Gly Pro Ala Ser Gly Pro
1 5 10 15
Thr Thr Asn Trp Gly Ile Pro Asn Ser Ala Ile Cys Ser Ser Gly Phe
20 25 30
Ser Gly Thr Thr Thr Pro Thr Val Pro Ser Val Ser Gly Asn Lys Pro
35 40 45
Val Thr Thr Ile Gln Gln Leu Ser Pro Ala Thr Ser Gly Ser Ala Ala

Substitute Sequence Listing_USSN 10587032_PP019482.007
55 60

Val Asp Leu Cys Thr Ile Gln Ala Val Ser Leu Leu Pro Gly Glu Pro
 65 70 75 80
 Pro Gln Lys Thr Pro Thr Gly Val Tyr Gly Pro Leu Pro Lys Gly Thr
 85 90 95
 Val Gly Leu Ile Leu Gly Arg Ser Ser Leu Asn Leu Lys Gly Val Gln
 100 105 110
 Ile His Thr Ser Val Val Asp Ser Asp Tyr Lys Gly Glu Ile Gln Leu
 115 120 125
 Val Ile Ser Ser Ser Ile Pro Trp Ser Ala Ser Pro Arg Asp Arg Ile
 130 135 140
 Ala Gln Leu Leu Leu Leu Pro Tyr Ile Lys Gly Gly Asn Ser Glu Ile
 145 150 155 160
 Lys Arg Ile Gly Gly Leu Gly Ser Thr Asp Pro Thr Gly Lys Ala Ala
 165 170 175
 Tyr Trp Ala Ser Gln Val Ser Glu Asn Arg Pro Val Cys Lys Ala Ile
 180 185 190
 Ile Gln Gly Lys Gln Phe Glu Gly Leu Val Asp Thr Gly Ala Asp Val
 195 200 205
 Ser Ile Ile Ala Leu Asn Gln Trp Pro Lys Asn Trp Pro Lys Gln Lys
 210 215 220
 Ala Val Thr Gly Leu Val Gly Ile Gly Thr Ala Ser Glu Val Tyr Gln
 225 230 235 240
 Ser Thr Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser Thr
 245 250 255
 Val Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg Asp
 260 265 270
 Leu Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Ser Tyr
 275 280 285
 Ser Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro Gly
 290 295 300
 Lys Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Ile Pro Val Glu Ala
 305 310 315 320
 Lys Ile Asn Gln Glu Arg Glu Gly Ile Gly Asn Pro Cys
 325 330

<210> 15
<211> 2896
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 15 atggcattaa aattccagtt gaggctaaaa taaaatcaaga aagagaagga ataggaaatc 60
cttgcttaggg gcggccactg tagagcctcc taaaacccata ccattaactt ggaaaacaga 120
aaaaccagtg tggtaaatc agtggccgct accaaaacaa aaactggagg ctttacattt 180
attaaatc aatccatgttgc aaaaatgtca tattttagcc tcgttttcac ctttggaaattc 240

Substitute Sequence Listing_USSN 10587032_PP019482.007

tccgtgttt	gtaattcaga	agaaatcagg	caaatggcg	atgttaactg	acttaaggc	300
tgttaacgcc	gtaattcaac	ccatggggcc	tctccaaccc	gggttgcctt	ctccggccat	360
gatcccaaaa	gattggcctt	taattataat	tgatctaaag	gattgctttt	ttaccatccc	420
tctggcagag	caggattgcg	aaaaatttgc	cttactata	ccagccataa	ataataaaga	480
accagccacc	aggtttcagt	ggaagtgtt	acctcaggg	atgcttaata	gtccaactat	540
ttgtcagact	ttttaggtc	gagctttca	accagttaga	gaaaagttt	cagactgtta	600
tattattcat	tgtattgtat	atattttatg	tgctgcagaa	acgaaagata	aattaattga	660
ctgttataca	tttctgcaag	cagaggttgc	caatgcttga	ctgcaatag	catctgataa	720
gatccaaacc	tctactcctt	ttcatttattt	aggatgcag	atagaaaata	aaaaatttaa	780
gccacaaaaa	atagaataaa	gaaaagacac	attaaaaaca	ctaatgatt	ttcaaaaaatt	840
actaggagat	attaatttga	ttcggccaac	tctaggcatt	cctactttag	ccatgtcaaa	900
tttgtctct	atcttaagag	gagactcaga	cttaaatagt	aaaagaatgt	taaccccaga	960
ggcaacaaaa	gaaatttaat	tagtggaga	aaaaatttcg	tcagcgc当地	taaatagaat	1020
agatccctta	gccccactcc	aactttgtat	tttgc当地	gcacattctc	caacaggcat	1080
cattattcaa	aatactgatc	ttgtggagtg	gtcattcctt	cctcacagta	cagttaaagac	1140
tttacatttgc	tacttggatc	aaatagctac	attaatcggt	cagacaagat	tacgaaataat	1200
aaaattatgt	ggaaatgacc	cagacaaaat	agttgtccct	ttaaccaagg	aacaaggtag	1260
acaaggcctt	atcaatttgc	gtgc当地	gattggctt	gctaattttt	tgggatttat	1320
tgataatcat	tacccaaaaa	caaagatctt	ccagttctt	aaatttgacta	cttggattct	1380
acctaatttgc	accagacgtg	aacctttaga	aatgcttca	acagtattt	ctgatgttcc	1440
cagcaatgga	aaagcagctt	acacaggacc	gaaagaacga	gtaatcaaa	ctccatataca	1500
atcggtctaa	agagcagat	tggttcag	cattacagtg	ttacaagatt	ttgaccaacc	1560
tatcaatatt	atatcgatt	ctgc当地	agtagcag	acaagggtt	ttgagacagc	1620
tctaatttttt	tatagcatgg	atgtcagtt	aaaccagctt	ttcaatttat	tacaacaaac	1680
tgtaagaaaa	agaaatttcc	cattttat	tacacatatt	cgagcacaca	ctaatttacc	1740
agggccttgc	actaaagcaa	atgaacaagg	tgacttactg	gtatcatctg	cactcataaa	1800
agcacaagaa	tttcatgttt	tgactcatgt	aaatgcagca	ggattaaaaa	acaaatttga	1860
tgtcacatgg	aaacaggcaa	aagatattgt	acaacatttc	accagtgtc	aagtcttaca	1920
cctgcccact	caagaggcag	gagttatcc	cagaggtctt	tgtcttaatg	cattatggca	1980
aatggatgtc	acgc当地	cttcatttgg	aagattatca	tatgttacg	taacagttga	2040
tacttattca	catttcatat	gggcaacttgc	ccaaacagga	gaaagtactt	cccatgtttaa	2100
aaaacatttgc	ttgtcttgc	ttgctgtat	gggagtttca	aaaaatttca	aaactgacaa	2160
tggaccagga	tattgtat	aagcttcc	aaaatttctt	agtcagttga	aaatttccaca	2220
tacaacagga	attccttata	attcccaagg	acaggccata	tttggaaagaa	ctaatacgtt	2280
actcaaaact	caatttgc	aacaaaaaga	agggggagac	agtaaggat	gtaccactcc	2340
tcagatgc当地	tttcatcttgc	cacttatac	ttttaatttt	ttaaacattt	atagaatca	2400
gactactact	tctgc当地	aacatcttac	ttgtaaaaag	aacagccac	atgaaggaaa	2460
actaatttgc	tggaaagata	ataaaaataa	gacatgggaa	atagggaaagg	tgataacgtt	2520
ggggagagat	tttgc当地	tttcaccagg	agaaaaatcag	tttgc当地	ggataaccac	2580
tagacatttgc	aatttgc当地	atgaaccat	cagagatgc当地	aagaaaagca	cctccgc当地	2640
gacggagaca	tcgcaatgc当地	gcaccgttgc当地	ctcacaagat	gaacaaaatg	gtgacgttgc当地	2700
aagaacagat	gaagttgc当地	tccaccaaga	aggcagagcc当地	gccaacttgc当地	gcacaactaa	2760
agaagctgac	gcagttatgc当地	acaaaatatc	tagagaacac	aaaggtgaca当地	caaacccttgc当地	2820
agagtatgtc	gcttgc当地	ttgatgttgc当地	tatcaatgttgc当地	ggttaagtctc当地	cctatgc当地	2880
caggagcagc	tgc当地					2896

<210> 16

<211> 2619

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 16

atgttaactg	acttaaggc	tgtaaacgcc	gtaattcaac	ccatggggcc	tctccaaccc	60
gggttgcctt	ctccggccat	gatcccaaaa	gattggcctt	taattataat	tgatctaaag	120
gattgctttt	ttaccatccc	tctggcagag	caggattgc当地	aaaaatttgc当地	cttactata	180
ccagccataa	ataataaaga	accagccacc	aggtttcagt	gaaaagtttgc当地	acctcaggg	240
atgcttaata	gtccaaactat	ttgtcagact	tttgc当地	gagcttctca	accagttaga	300
gaaaagtttgc当地	cagactgttgc当地	tattttat	tgtattgtat	atattttatg	tgctgc当地	360
acgaaagata	aattaatttgc当地	ctgttataca	tttctgcaag	cagaggttgc当地	caatgcttgc当地	420
ctggcaatag	catctgatata	gatccaaacc	tctactcctt	ttcatttattt	agggatgc当地	480
atagaaaataa	gaaaaatttgc当地	gccacaaa	atagaatataa	gaaaagacac	ataaaaacca	540
cttaatgttgc当地	ttcaaaaaatttgc当地	actaggat	attaatttgc当地	ttcggccacc	tctaggcatt	600
cctacttata	ccatgtcaaa	tttgc当地	atcttaagag	gagactcaga	cttaaatagt	660

Substitute Sequence Listing_USSN 10587032_PP019482.007

aaaagaatgt	taaccccaga	ggcaacaaaa	gaaattaaat	tagtggaga	aaaaatccag	720
tcagcgc当地	taaatagaat	agatccctt	gccccactcc	aactttgat	tttgccact	780
gcacattctc	caacaggcat	cattattcaa	aatactgatc	ttgtggagt	gtcattcctt	840
cctcacagta	cagttagac	tttacattg	tacttggatc	aatagctac	attaatcggt	900
cagacaagat	taogaataat	aaaattatgt	gggaatgacc	cagacaaaat	agttgtccct	960
ttaaccagg	aacaagttag	acaaggctt	atcaattctg	gtgcattggaa	gattggctt	1020
gctaattttt	tgggattat	tgataatcat	tacccaaaaa	caaagatctt	ccagttctta	1080
aaattgacta	cttggattt	acctaaaatt	accagacgt	aacctttaga	aatgctcta	1140
acagtattt	ctgatggttc	cagcaatgga	aaagcagctt	acacaggacc	gaaagaacga	1200
gtaatcaaaa	ctccatata	atcggtcaa	agagcagat	tgggtcagt	cattacagt	1260
ttacaagatt	ttgaccaacc	tatcaatatt	atatcgatt	ctgcataatgt	agtacaggct	1320
acaaggatg	ttgagacagc	tctaattaaa	tatagcatgg	atgatcgat	aaaccagcta	1380
ttcaatttat	tacaacaaac	tgtaagaaaa	agaaatttcc	cattttat	tacacatatt	1440
cgagcacaca	ctaatttacc	agggcctt	actaaagcaa	atgaacaagc	tgacttactg	1500
gtatcatctg	cactcataaa	agcacaagaa	cttcatgtt	tgactcatgt	aatgcagca	1560
ggattaaaaa	acaaatttga	tgtcacatgg	aaacaggcca	aagatattgt	acaacattgc	1620
acccagtgtc	aagtcttaca	cctgcccact	caagaggcag	gagttatcc	cagaggtctg	1680
tgtcctaattg	cattatggca	aatggatgtc	acgcatgtac	cttcatttgg	aagattatca	1740
tatgttcacg	taacagttga	tacttattca	catttcat	ggcaacttgc	ccaaacacgga	1800
gaaagtactt	cccatgtttaa	aaaacatttta	ttgtcttgg	ttgtgtat	gggagttcca	1860
gaaaaaaaaatca	aaactgacaa	tggaccagga	tattttagta	aagtttcca	aaaatttta	1920
agtcaatgttga	aaatttccaca	tacaacagga	atttccatttta	atttccaagg	acaggccata	1980
gttgaagaa	ctaatagaac	actcaaaact	caatttagtt	aaaaaaaaga	agggggagac	2040
agtaaggagt	gtaccactcc	tcagatgca	cttaatctag	cactctatac	tttaattttt	2100
ttaaacattt	atagaaatca	gactactact	tctgcagaac	aacatcttac	ttgtaaaaag	2160
aacagccccac	atgaaggaaa	actaatttgg	tggaaagata	gtaaaaataa	gacatggaa	2220
atagggaaagg	tgataacgt	ggggagaggt	tttgcttgc	tttcaccagg	agaaaaatccag	2280
cttcctgttt	ggataccac	tagacatttgc	aagttctaca	atgaacccat	cagagatgca	2340
aagaaaaagca	cctccgcgg	gacggagaca	tcgcaatcga	gcaccgttga	ctcacaagat	2400
gaacaaaatg	gtgacgtcag	aagaacagat	gaagttgcca	tccaccaaga	aggcagagcc	2460
gccaacttgg	gcacaactaa	agaagctgac	gcagttagct	acaaaatatc	tagagaacac	2520
aaaggtgaca	caaacccccag	agagtatgt	gcttgcagcc	ttgtatgattt	tatcaatgtt	2580
ggtaagtctc	cctatgcctg	caggagcagc	tgcaatgtt			2619

<210> 17

<211> 2671

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 17

atggcattaa	agtcccaact	gaggctgaaa	aaaatcaaaa	aaagaaaaagg	aataggcat	60
cctttttaga	agcggtcact	gtagagcctc	aaaacccat	tccattaatt	tggggggaaa	120
aaaaaaactg	tatgtttaat	cagtagccgc	ttccaaaaca	aaaactggag	gttttacact	180
tattagcaaa	gaaacagtta	aaaaaaggac	atattgagcc	ttcattttcg	ctttggatt	240
ctcctgttt	taattcagaa	aaaatccggc	agatggcgta	tgcttaactga	cttaagagcc	300
attaatgcca	taattcaacc	catggggct	ctcccatccc	ggttgcctc	tccagccatg	360
gtccccctta	attataattt	atctgaagga	ttgcttttt	accattcctc	tggcaaaaaga	420
ggattttgaa	aaatttgctt	ttactatacc	agcctaaata	ataaaagaacc	agccaccagg	480
tttcagtgg	aagtatttgc	tcagggatg	cttaataatt	caactatttgc	tcagactttc	540
atagctcaag	ctctgcaacc	agtttagagac	aagttttcag	actgttat	cgttcattat	600
gttgatattt	tgtgtctgc	agaaacagaga	gacaaattaa	ttgaccgtt	cacatttctc	660
agacagaggt	tgccaaacgcg	ggactgacaa	tagcatctga	taagattcaa	acctctcc	720
cttccattt	cttgggaaatg	caggttagagg	aaaggaaaaat	taaaccacaa	aaaatagaaa	780
taagaaaaaga	cacattaaaa	acattaaatg	agtttcaaaa	tttggtagga	gatactaatt	840
ggattcggag	atattaaattt	gatttggcca	actcttaggca	ttcctactt	tgccatgtca	900
attttgttct	ctttcttaag	aggggacttg	gaatttataa	gtgaaagaat	gttacccca	960
gaggcaacta	aagaaaattaa	attaattgaa	aaaaaaaatt	cggtcagcac	aagtaaatag	1020
gatcacttgg	ccccactcca	aattttgatt	tttggtactg	cacattctct	aacagccatc	1080
attgttcaaa	acacagatct	tgtggattgg	tccttccttc	ctcatagta	aattaagact	1140
tttacattgt	acttggatca	aatggctaca	ttaatttgc	agggaaagatt	acgaataata	1200
acattgtgtt	gaaaatgaccc	agataaaaatc	actgttccctt	tcaacaaagca	acaagttaga	1260
caagccctta	tcaatgttgc	tgcattggcag	atttgcatttgc	ctaattttct	ggaattattt	1320
gataatcatt	acccaaaaac	aaaaatcttc	cagttctttaa	aatttgactac	ttggattctt	1380

Substitute Sequence Listing_USSN 10587032_PP019482.007

cctaaaatta	ccagacgtga	acctttagaa	aatgctctaa	cagtattac	tgatggttcc	1440
agcaatggaa	aagcggctta	cacaggccg	aaagaacgag	taatcaaaaac	tccgtatcaa	1500
tcagctcaaa	gagcagagg	gttgcagtc	attacagtgt	tacaagattt	tgaccaacct	1560
atcaatatta	tatcgat	tgcatatgt	gtacaggcta	caaggatgt	tgagacagct	1620
ctaattaaat	atagcacgga	cgatcattt	aaccagctat	tcaattttat	acaacaaact	1680
gtaagaaaaa	gaaatttccc	attttatatt	actcatattc	gagcacacac	taatttacca	1740
gggccttga	ctaaagcaa	tgaacaagct	gacttactgg	tatcatctgc	attcataaaa	1800
gcacaagaac	ttcttgctt	gactcatgt	aatgcagcag	gataaaaaaaaa	caaatttgat	1860
gtcacatgga	aacaggcaa	agatattgt	caacattgca	cccagtgtca	agtcttacac	1920
ctgtccactc	aagaggcagg	agttatccc	agaggtctgt	gtcctaattgc	gttatggcaa	1980
atggatggca	cgcattttcc	ttcatttgg	agattatcat	atgttcatgt	aacagttgat	2040
acttattcac	atttcatatg	ggcaacttgc	caaacaggag	aaagtacttc	ccatgttaaa	2100
aaacatttat	tatcttgtt	tgctgtatg	ggagttccag	aaaaaatcaa	aactgacaat	2160
ggaccaggat	attgttagtaa	agctttccaa	aaattcttaa	gtcagtggaa	aatttcacat	2220
acaacaggaa	ttccttataa	ttcccaagga	caggccatag	ttgaaagaac	taatagaaca	2280
ctcaaaactc	aattagttaa	acaaaaagaa	gggggagaca	gtaaggagtg	taccactcct	2340
cagatgcac	ttatcttagc	actctatact	ttaaattttt	taaacattta	tagaaatcag	2400
actactactt	ctgcaaaaac	acatcttact	gttaaaaggc	acagcccaca	tgaagggaaa	2460
ctaatttgg	gaaaaagataa	taaaaataag	acatggggaa	taggaagg	gataacgtgg	2520
gggagagggt	ttgcttgtt	ttcaccagga	aaaaatcagc	ttcctgtttt	gataccact	2580
agacatttga	agttctacaa	tgaacccatc	ggagatgca	agaaaaggc	ctccacagag	2640
atggtaaccc	cagtcacatg	gatggataat	C			2671

<210> 18

<211> 4086

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 18

atggggcctc	tccaacccgg	gttgcctct	ccggccatga	tccaaaaga	ttggccttta	60
attataattg	atctaaagg	ttgctttt	accatccctc	tggcagagca	ggatttgaa	120
aaatttgcct	ttactatacc	agccataaaat	aataaagaac	cagccaccag	gtttcaigtgg	180
aaagtgttac	ctcagggat	gcttaatagt	ccaaactattt	gtcagacttt	tgttaggtcga	240
gctctcaac	cagtgagaga	aaagtttca	gactgttata	ttattcatta	tattgtatgat	300
attttatgt	ctgcagaaac	gaaagataaa	ttaatttact	gttatacatt	tctgcaagca	360
gaggttgcca	atgtctggact	ggcaatagca	tccgataaga	tccaaacctc	tactcccttt	420
cattattn	ggatgcagat	agaaaaataga	aaaattaaagc	cacaaaaat	agaaataaga	480
aaagacacat	taaaaacact	aatgatttt	caaaaattac	tagagat	taatttgatt	540
cggccaaactc	taggcatttcc	tactttagcc	atgtcaattt	tgttctctat	cttaagagga	600
gactcagact	taaaatagtc	aagaatatta	accccagagg	caacaaaaga	aattaaatta	660
gtggaaagaaa	aaatttcagtc	agcgcataata	aatagaatag	atccctttagc	cccactccaa	720
cttttgcattt	ttgccactgc	acattctca	acaggcatca	ttattcaaaa	tactgtatctt	780
gtggagtgtt	cattccttcc	tcacagtaca	gttaagactt	ttacattgt	cttggatcaa	840
atagctacat	taatcggtca	gacaagatta	cgaataacaa	aattatgtgg	aatgacccca	900
gacaaaatag	ttgtcccttt	aaccaaggaa	caagtttagac	aaggctttat	caattctgg	960
gcatggcaga	ttggtcttg	taattttgt	ggacttattt	ataatcatta	cccaaaaaca	1020
aagatcttcc	agttcttaaa	attgactact	tggattctac	ctaaaattac	cagacgtgaa	1080
cctttagaaa	atgctctaa	agtattttact	gatggttcca	gcaatggaaa	agcagcttac	1140
acagggccga	aagaacgagt	aatcaaaaact	ccatatcaat	cggctcaag	agacgagttg	1200
gttgcagtca	ttacagtgtt	acaagatttt	gaccaaccta	tcaatattat	atcagattct	1260
gcatatgttag	tacaggctac	aaggatgtt	gagacagctc	taattaaata	tagcatggat	1320
gatcagttaa	accagctatt	caatttatta	caacaaactg	taagaaaaag	aaatttccca	1380
ttttatata	cttatatttc	agcacacact	aatttaccag	ggccttgc	taaagcaa	1440
gaacaagctg	acttactgg	atcatctgca	ctcataaaag	cacaagaact	tcatgcttgc	1500
actcatgtaa	atgcagcagg	attaaaaaaac	aaatttgcgt	tcacatggaa	acaggcaaaa	1560
gatattgtac	aacattgcac	ccagtgtaa	gttcttacacc	tgcccactca	agaggcagga	1620
gttaatccca	gaggtctgt	tcctaatgt	ttatggcaaa	tggatgtc	gcatgtac	1680
tcatttggaa	gattatcata	ttttcatgt	acagttgtat	tttattcaca	tttcatatgg	1740
gcaacttgcc	aaacaggaga	aagtacttcc	catgttaaaa	aacatttt	gtcttgc	1800
gctgtatgg	gagttccaga	aaaaatcaa	actgacaatg	gaccaggata	ttgttagtaaa	1860
gctttccaaa	aattcttaag	tcagtgaaa	atttcacata	caacaggaa	tccttataat	1920
tcccaaggac	aggccatagt	tgaagaact	aatagaacac	tcaaaaactca	attagttaaa	1980
caaaaagaag	ggggagacag	taaggagtgt	accactcctc	agatgcaact	taatctagca	2040

Substitute Sequence Listing_USSN 10587032_PP019482.007

ctctatactt	taaattttt	aaacatttat	agaaaatcaga	ctactacttc	tgcagaacaa	2100
catctactg	gtaaaaaaa	cagcccacat	gaaggaaaac	taatttggtg	gaaagataat	2160
aaaaataaga	catggaaat	agggaaaggtg	ataacgtgg	ggagaggttt	tgcttggtt	2220
tcaccaggag	aaaatcagct	tcctgttgg	ttacccacta	gacattgaa	gttctacaat	2280
gaaccatcg	gagatgcaa	gaaaagggcc	tccacggaga	tggtaacacc	agtcatcg	2340
atggataatc	ctatagaagt	atatgttaat	gatagtatat	gggtacctgg	ccccatagat	2400
gatcgctgcc	ctgccaaacc	tgaggaagaa	gggatgatga	taaatatttc	cattgggtat	2460
cgttatcctc	ctatggcct	agggagagca	ccaggatgtt	taatgcctgc	agtccaaaat	2520
tggttggtag	aagtacctac	tgtcagtccc	atcagtagat	tcacttatca	catggtaagc	2580
gggatgtcac	tcaggccacg	ggtaaattat	ttacaagact	tttcttatca	aagatcatta	2640
aaatttagac	ctaaaggaa	acctgcccc	aaggaaattc	ccaaagaatc	aaaaaataca	2700
gaagtttag	tttggaaaga	atgtgtggcc	aatagtgcgg	tgatattata	aaacaatgaa	2760
tttggacta	ttatagattg	ggcacctcga	ggtcaattct	accacaattg	ctcaggacaa	2820
actcagtcgt	gtccaagtgc	acaagtgagt	ccagctttg	atagcactt	aacagaaagt	2880
ttagacaaac	ataagcataa	aaaattgcag	tctttctacc	cttggaaatg	gggagaaaaa	2940
ggaatctcta	ccccaaagacc	aaaaatagta	agtccctttt	ctgtcctgta	acatccagaa	3000
ttatggaggc	ttactgtggc	ctcacaccac	attagaattt	ggctctggaaa	tcaaacttta	3060
gaaacaaagag	attgtaaagcc	attttatact	gtcgaccata	atttcagtt	aacagttcct	3120
ttacaagtt	gcgttaaagcc	cccttatatg	ctagttgttag	gaaatatagt	tattaacca	3180
gactccaga	ctataacctg	tgaaaattgt	agattgccta	cttgcattga	ttcaactttt	3240
aattggcaac	accgtattct	gctgtgaga	gcaagagagg	gcgtgtggat	ccctgtgtcc	3300
atggaccgac	cgtggggagc	ctcaccatcc	gtccatattt	tgactgaagt	attaaaagg	3360
gttttaataa	gatccaaaag	attcattttt	actttaattt	cagtgattat	gggattaatt	3420
gcagtcacag	ctacggctgc	tgttagcagga	gttgcattgc	actcttctgt	tcagtcagta	3480
aactttgtta	atgattggca	aaagaattct	acaagattgt	ggaattcaca	atctagtatt	3540
gatcaaaaat	tggcaaatca	aattaatgtat	cttagacaaa	ctgtcatttg	gatggagac	3600
agactcatga	gcttagaaca	tcgtttccag	ttacaatgtt	actggaatac	gtcagatttt	3660
tgtattacac	cccaaattta	taatgagtct	gagcatca	gggacatgg	tagacccat	3720
ctacagggaa	gagaagataa	tctcaatttt	gacatttca	aattaaaaga	acaaattttc	3780
gaagcatcaa	aagccattt	aaatttggtg	ccaggaactg	aggcaattgc	aggagttgct	3840
gatggcctcg	caaatcttaa	ccctgtca	tgggtaaga	ccattggaa	tacatcgatt	3900
ataaatctca	tattaatctt	tgtgtgcctg	tttgtctgt	tgttagtctg	caggtgtacc	3960
caacagctcc	gaagagacag	cgaccatcga	gaacgggcca	tgtgacgat	ggcggtttt	4020
tcgaaaagaa	aagggggaaa	tgtggggaaa	agcaagagag	atcaaattgt	tactgtgtct	4080
						4086

<210> 19

<211> 872

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 19

Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val Ile Gln Pro Met Gly
1 5 10 15

Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile Pro Lys Asp Trp
20 25 30

Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe Phe Thr Ile Pro Leu
35 40 45

Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr Ile Pro Ala Ile Asn
50 55 60

Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys Val Leu Pro Gln Gly
65 70 75 80

Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe Val Gly Arg Ala Leu
85 90 95

Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr Ile Ile His Cys Ile
100 105 110

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys Leu Ile Asp Cys
115 120 125

Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly Leu Ala Ile Ala
130 135 140

Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr Leu Gly Met Gln
145 150 155 160

Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu Ile Arg Lys Asp
165 170 175

Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu Gly Asp Ile Asn
180 185 190

Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala Met Ser Asn Leu
195 200 205

Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser Lys Arg Met Leu
210 215 220

Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu Glu Lys Ile Gln
225 230 235 240

Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro Leu Gln Leu Leu
245 250 255

Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile Ile Gln Asn Thr
260 265 270

Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr Val Lys Thr Phe
275 280 285

Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly Gln Thr Arg Leu
290 295 300

Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp Lys Ile Val Val Pro
305 310 315 320

Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn Ser Gly Ala Trp
325 330 335

Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile Asp Asn His Tyr Pro
340 345 350

Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr Trp Ile Leu Pro
355 360 365

Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu Thr Val Phe Thr
370 375 380

Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly Pro Lys Glu Arg
385 390 395 400

Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Ala Glu Leu Val Ala
405 410 415

Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile Asn Ile Ile Ser
420 425 430

Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val Glu Thr Ala Leu
435 440 445

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu Phe Asn Leu Leu
450 455 460

Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr Ile Thr His Ile
465 470 475 480

Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys Ala Asn Glu Gln
485 490 495

Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala Gln Glu Leu His
500 505 510

Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn Lys Phe Asp Val
515 520 525

Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys Thr Gln Cys Gln
530 535 540

Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn Pro Arg Gly Leu
545 550 555 560

Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His Val Pro Ser Phe
565 570 575

Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr Tyr Ser His Phe
580 585 590

Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser His Val Lys Lys
595 600 605

His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro Glu Lys Ile Lys
610 615 620

Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe Gln Lys Phe Leu
625 630 635 640

Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro Tyr Asn Ser Gln
645 650 655

Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu Lys Thr Gln Leu
660 665 670

Val Lys Gln Lys Glu Gly Asp Ser Lys Glu Cys Thr Thr Pro Gln
675 680 685

Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe Leu Asn Ile Tyr
690 695 700

Arg Asn Gln Thr Thr Ser Ala Glu Gln His Leu Thr Gly Lys Lys
705 710 715 720

Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp Ser Lys Asn
725 730 735

Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly Arg Gly Phe Ala
740 745 750

Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Ile Pro Thr Arg
755 760 765

His Leu Lys Phe Tyr Asn Glu Pro Ile Arg Asp Ala Lys Lys Ser Thr
770 775 780

Substitute Sequence Listing_USSN 10587032_PP019482.007
Ser Ala Glu Thr Glu Thr Ser Gln Ser Ser Thr Val Asp Ser Gln Asp
785 790 795 800

Glu Gln Asn Gly Asp Val Arg Arg Thr Asp Glu Val Ala Ile His Gln
805 810 815

Glu Gly Arg Ala Ala Asn Leu Gly Thr Thr Lys Glu Ala Asp Ala Val
820 825 830

Ser Tyr Lys Ile Ser Arg Glu His Lys Gly Asp Thr Asn Pro Arg Glu
835 840 845

Tyr Ala Ala Cys Ser Leu Asp Asp Cys Ile Asn Gly Gly Lys Ser Pro
850 855 860

Tyr Ala Cys Arg Ser Ser Cys Ser
865 870

<210> 20

<211> 1361

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<220>

<221> SITE

<222> 917

<223> Xaa is any amino acid

<400> 20
Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile Pro Lys
1 5 10 15

Asp Trp Pro Leu Ile Ile Asp Leu Lys Asp Cys Phe Phe Thr Ile
20 25 30

Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr Ile Pro Ala
35 40 45

Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys Val Leu Pro
50 55 60

Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe Val Gly Arg
65 70 75 80

Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr Ile Ile His
85 90 95

Tyr Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys Leu Ile
100 105 110

Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly Leu Ala
115 120 125

Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr Leu Gly
130 135 140

Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu Ile Arg
145 150 155 160

Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu Gly Asp
165 170 175

Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala Met Ser

Substitute Sequence Listing_USSN 10587032_PP019482.007
180 185 190

Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser Gln Arg
195 200 205

Ile Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu Glu Lys
210 215 220

Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro Leu Gln
225 230 235 240

Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile Ile Gln
245 250 255

Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr Val Lys
260 265 270

Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly Gln Thr
275 280 285

Arg Leu Arg Ile Thr Lys Leu Cys Gly Asn Asp Pro Asp Lys Ile Val
290 295 300

Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn Ser Gly
305 310 315 320

Ala Trp Gln Ile Gly Leu Ala Asn Phe Val Gly Leu Ile Asp Asn His
325 330 335

Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr Trp Ile
340 345 350

Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu Thr Val
355 360 365

Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly Pro Lys
370 375 380

Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Asp Glu Leu
385 390 395 400

Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile Asn Ile
405 410 415

Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val Glu Thr
420 425 430

Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu Phe Asn
435 440 445

Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr Ile Thr
450 455 460

Tyr Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys Ala Asn
465 470 475 480

Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala Gln Glu
485 490 495

Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn Lys Phe
500 505 510

Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys Thr Gln
Page 22

515 Substitute Sequence Listing_USSN 10587032_PP019482.007
520 525

Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn Pro Arg
530 535 540

Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His Val Pro
545 550 555 560

Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr Tyr Ser
565 570 575

His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser His Val
580 585 590

Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro Glu Lys
595 600 605

Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe Gln Lys
610 615 620

Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro Tyr Asn
625 630 635 640

Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu Lys Thr
645 650 655

Gln Leu Val Lys Gln Lys Glu Gly Asp Ser Lys Glu Cys Thr Thr
660 665 670

Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe Leu Asn
675 680 685

Ile Tyr Arg Asn Gln Thr Thr Ser Ala Glu Gln His Leu Thr Gly
690 695 700

Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp Asn
705 710 715 720

Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly Arg Gly
725 730 735

Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Leu Pro
740 745 750

Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Gly Asp Ala Lys Lys
755 760 765

Arg Ala Ser Thr Glu Met Val Thr Pro Val Thr Trp Met Asp Asn Pro
770 775 780

Ile Glu Val Tyr Val Asn Asp Ser Ile Trp Val Pro Gly Pro Ile Asp
785 790 795 800

Asp Arg Cys Pro Ala Lys Pro Glu Glu Gly Met Met Ile Asn Ile
805 810 815

Ser Ile Gly Tyr Arg Tyr Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly
820 825 830

Cys Leu Met Pro Ala Val Gln Asn Trp Leu Val Glu Val Pro Thr Val
835 840 845

Ser Pro Ile Ser Arg Phe Thr Tyr His Met Val Ser Gly Met Ser Leu

850

Substitute Sequence Listing_USSN 10587032_PP019482.007
855 860

Arg Pro Arg Val Asn Tyr Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu
865 870 875 880

Lys Phe Arg Pro Lys Gly Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu
885 890 895

Ser Lys Asn Thr Glu Val Leu Val Trp Glu Glu Cys Val Ala Asn Ser
900 905 910

Ala Val Ile Leu Xaa Asn Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala
915 920 925

Pro Arg Gly Gln Phe Tyr His Asn Cys Ser Gly Gln Thr Gln Ser Cys
930 935 940

Pro Ser Ala Gln Val Ser Pro Ala Val Asp Ser Asp Leu Thr Glu Ser
945 950 955 960

Leu Asp Lys His Lys His Lys Leu Gln Ser Phe Tyr Pro Trp Glu
965 970 975

Trp Gly Glu Lys Gly Ile Ser Thr Pro Arg Pro Lys Ile Val Ser Pro
980 985 990

Val Ser Gly Pro Glu His Pro Glu Leu Trp Arg Leu Thr Val Ala Ser
995 1000 1005

His His Ile Arg Ile Trp Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp
1010 1015 1020

Cys Lys Pro Phe Tyr Thr Val Asp Leu Asn Ser Ser Leu Thr Val Pro
1025 1030 1035 1040

Leu Gln Ser Cys Val Lys Pro Pro Tyr Met Leu Val Val Gly Asn Ile
1045 1050 1055

Val Ile Lys Pro Asp Ser Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu
1060 1065 1070

Leu Thr Cys Ile Asp Ser Thr Phe Asn Trp Gln His Arg Ile Leu Leu
1075 1080 1085

Val Arg Ala Arg Glu Gly Val Trp Ile Pro Val Ser Met Asp Arg Pro
1090 1095 1100

Trp Glu Ala Ser Pro Ser Val His Ile Leu Thr Glu Val Leu Lys Gly
1105 1110 1115 1120

Val Leu Asn Arg Ser Lys Arg Phe Ile Phe Thr Leu Ile Ala Val Ile
1125 1130 1135

Met Gly Leu Ile Ala Val Thr Ala Ala Val Ala Gly Val Ala
1140 1145 1150

Leu His Ser Ser Val Gln Ser Val Asn Phe Val Asn Asp Trp Gln Lys
1155 1160 1165

Asn Ser Thr Arg Leu Trp Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu
1170 1175 1180

Ala Asn Gln Ile Asn Asp Leu Arg Gln Thr Val Ile Trp Met Gly Asp

1185 Substitute Sequence Listing_USSN 10587032_PP019482.007
 1190 1195 1200

Arg Leu Met Ser Leu Glu His Arg Phe Gln Leu Gln Cys Asp Trp Asn
1205 1210 1215

Thr Ser Asp Phe Cys Ile Thr Pro Gln Ile Tyr Asn Glu Ser Glu His
1220 1225 1230

His Trp Asp Met Val Arg Arg His Leu Gln Gly Arg Glu Asp Asn Leu
1235 1240 1245

Thr Leu Asp Ile Ser Lys Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys
1250 1255 1260

Ala His Leu Asn Leu Val Pro Gly Thr Glu Ala Ile Ala Gly Val Ala
1265 1270 1275 1280

Asp Gly Leu Ala Asn Leu Asn Pro Val Thr Trp Val Lys Thr Ile Gly
1285 1290 1295

Ser Thr Ser Ile Ile Asn Leu Ile Leu Ile Leu Val Cys Leu Phe Cys
1300 1305 1310

Leu Leu Leu Val Cys Arg Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp
1315 1320 1325

His Arg Glu Arg Ala Met Met Thr Met Ala Val Leu Ser Lys Arg Lys
1330 1335 1340

Gly Gly Asn Val Gly Lys Ser Lys Arg Asp Gln Ile Val Thr Val Ser
1345 1350 1355 1360

Val

<210> 21
<211> 956
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 21
Asn Lys Ser Arg Lys Arg Arg Asn Arg Glu Ser Leu Leu Gly Ala Ala
1 5 10 15

Thr Val Glu Pro Pro Lys Pro Ile Pro Leu Thr Trp Lys Thr Glu Lys
20 25 30

Pro Val Trp Val Asn Gln Trp Pro Leu Pro Lys Gln Lys Leu Glu Ala
35 40 45

Leu His Leu Leu Ala Asn Glu Gln Leu Glu Lys Gly His Ile Glu Pro
50 55 60

Ser Phe Ser Pro Trp Asn Ser Pro Val Phe Val Ile Gln Lys Lys Ser
65 70 75 80

Gly Lys Trp Arg Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val Ile
85 90 95

Gln Pro Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile
100 105 110

Pro Lys Asp Trp Pro Leu Ile Ile Asp Leu Lys Asp Cys Phe Phe
Page 25

115 Substitute Sequence Listing_USSN 10587032_PP019482.007
 120 125

Thr Ile Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr Ile
130 135 140

Pro Ala Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys Val
145 150 155 160

Leu Pro Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe Val
165 170 175

Gly Arg Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr Ile
180 185 190

Ile His Cys Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys
195 200 205

Leu Ile Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly
210 215 220

Leu Ala Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr
225 230 235 240

Leu Gly Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu
245 250 255

Ile Arg Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu
260 265 270

Gly Asp Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala
275 280 285

Met Ser Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser
290 295 300

Lys Arg Met Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu
305 310 315 320

Glu Lys Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro
325 330 335

Leu Gln Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile
340 345 350

Ile Gln Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr
355 360 365

Val Lys Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly
370 375 380

Gln Thr Arg Leu Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp Lys
385 390 395 400

Ile Val Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn
405 410 415

Ser Gly Ala Trp Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile Asp
420 425 430

Asn His Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr
435 440 445

Trp Ile Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu

450

Substitute Sequence Listing_USSN 10587032_PP019482.007
455 460Thr Val Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly
465 470 475 480Pro Lys Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Ala
485 490 495Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile
500 505 510Asn Ile Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val
515 520 525Glu Thr Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu
530 535 540Phe Asn Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr
545 550 555 560Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys
565 570 575Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala
580 585 590Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn
595 600 605Lys Phe Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys
610 615 620Thr Gln Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn
625 630 635 640Pro Arg Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His
645 650 655Val Pro Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr
660 665 670Tyr Ser His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser
675 680 685His Val Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro
690 695 700Glu Lys Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe
705 710 715 720Gln Lys Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro
725 730 735Tyr Asn Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu
740 745 750Lys Thr Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys
755 760 765Thr Thr Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe
770 775 780Leu Asn Ile Tyr Arg Asn Gln Thr Thr Ser Ala Glu Gln His Leu
Page 27

Substitute Sequence Listing_USSN 10587032_PP019482.007
 785 790 795 800

Thr Gly Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys
 805 810 815

Asp Asn Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly
 820 825 830

Arg Gly Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp
 835 840 845

Ile Pro Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Arg Asp Ala
 850 855 860

Lys Lys Ser Thr Ser Ala Glu Thr Glu Thr Ser Gln Ser Ser Thr Val
 865 870 875 880

Asp Ser Gln Asp Glu Gln Asn Gly Asp Val Arg Arg Thr Asp Glu Val
 885 890 895

Ala Ile His Gln Glu Gly Arg Ala Ala Asn Leu Gly Thr Thr Lys Glu
 900 905 910

Ala Asp Ala Val Ser Tyr Lys Ile Ser Arg Glu His Lys Gly Asp Thr
 915 920 925

Asn Pro Arg Glu Tyr Ala Ala Cys Ser Leu Asp Asp Cys Ile Asn Gly
 930 935 940

Gly Lys Ser Pro Tyr Ala Cys Arg Ser Ser Cys Ser
 945 950 955

<210> 22
 <211> 2000

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 22

atgaacccat	cagagatgca	aagaaaagca	cctccgcgga	gacggagaca	tcgcaatcga	60
gcaccgttga	ctcacaagat	gaacaaaaatg	gtgacgtcag	aagaacagat	gaagttgcca	120
tccaccaaga	aggcagagcc	gccaaacttgg	gcacaactaa	agaagctgac	gcagtttagct	180
acaaaataatc	tagagaacac	aaagggtgaca	caaaccggcag	agagtatgtct	gcttgcagcc	240
tttatgtattt	tatcaatgtt	ggtaagtctc	cctatgcctg	caggagcagc	tgcagctaacc	300
tatacctact	gggcctatgt	gcctttcccg	cccttaattt	gggcagtcac	atggatggat	360
aatcctacag	aagtatatgt	taatgtatgt	gtatgggtac	ctggccccat	agatgatcgc	420
tgccctgcca	aacctgagga	agaagggatg	atgataaata	tttccatttg	gtatcattat	480
cctccttattt	gccttagggag	agcaccagga	tgtttaatgc	ctgcagtcac	aaattgttg	540
gtagaagtac	ctactgtcag	tcccatttgt	agattcactt	atcacatgtt	aagcggatg	600
tcactcaggc	cacggtaaaa	ttatttacaa	gacttttctt	atcaaagatc	attaaaattt	660
agacctaaag	ggaaaccttg	ccccaaaggaa	attcccaaag	aataaaaaaaa	tacagaagtt	720
tttagtttggg	aagaatgtgt	ggccaaatagt	gcggtgatata	tacaaaacaa	tgaattcgg	780
actattatag	attgggcacc	tcgaggtcaa	ttctaccaca	attgctcagg	acaaaactcag	840
tcgtgtccaa	gtgcacaagt	gagtccagct	gttgatagcg	acttaacaga	aagtttagac	900
aaacataagc	ataaaaaaattt	gcagtcttc	tacccttgg	aatggggaga	aaaaggaatc	960
tctaccccaa	gacaaaaaat	agtaagtcc	gtttctggc	ctgaacatcc	agaatttatgg	1020
aggcttactg	tggcctcaca	ccacatttga	atttggctcg	gaaatcaaacc	tttagaaaca	1080
agagatcgta	agccatttta	tactattgac	ctgaatttca	gtctaacagt	tcctttacaa	1140
agttgcgtaa	agccccctta	tatgttagtt	gttagaaata	tagttattaa	accagactcc	1200
cagactataa	cctgtaaaaa	ttgttagattt	cttacttgc	ttgattcaac	tttaatttgg	1260
caacaccgtt	ttctgctgg	gagagcaaga	gagggcggt	ggatccctgt	gtccatggac	1320
cgaccgtggg	aggcctcgcc	atccgtccat	attttgactg	aagttttaaa	aggtgtttt	1380
aatagatcca	aaagattcat	ttttacttta	attgcagtg	ttatgggatt	aattgcagtc	1440
acagctacgg	ctgctgttagc	aggagttgca	ttgcactt	ctgttcagtc	agtaaacttt	1500

Substitute Sequence Listing_USSN 10587032_PP019482.007

gttaatgatt	ggcaaaaaaa	ttctacaaga	ttgtggatt	cacaatctag	tattgatcaa	1560
aaatggcaa	atcaaattaa	tgatctaga	caaactgtca	tttggatggg	agacagactc	1620
at gagcttag	a acatcg tt	ccagttacaa	tgtgactgga	atacgtcaga	ttttgtatt	1680
acacccaaa	tttataatga	gtctgagcat	cactggac a	tggttagacg	ccatctacag	1740
ggaagagaag	ataatctcac	tttagacatt	tccaaattaa	aagaacaaat	tttcgaagca	1800
tcaa aagccc	at taaaattt	ggtgccagga	actgaggca a	ttgcaggagt	tgctgtatgc	1860
ctcgcaa atc	ttaaccctgt	cactgggtt	aagaccattg	gaagtactac	gattataaat	1920
ctcatattaa	tccttgtgt	cctgtttgt	ctgttgttag	tctgcagggt	tacccaacag	1980
ctccgaagag	acagcgacca					2000

<210> 23

<211> 2085

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 23

atgcaaagaa	a a g c a c t c c	g c g g a g a c g g	a g a c a t c g c a	a t c g a g c a c c	g t t g a c t c a c	60
a a g a t g a a c a	a a a t g g t g a c	g t c a g a a g a a	c a g a t g a a g t	t g c c a t c c a c	c a a g a a g g c a	120
g a g c c c c a a	c t t g g g c a c a	a c t a a a g a a g	c t g a c g c a g t	t a g c t a c a a a	a t a t c t a g a g	180
a a c a c a a a g g	t g a c a c a a a c	c c c a g a g a g t	a t g c t g c t t g	c a g c c t t g a t	g a t t g t a t c a	240
a t g g t g t a a a	g t c t c c t t a t	g c c t g c a g g a	g c a g c t g c a g	c t a a c t a t a c a	c t a c t g g g c c	300
t a t g t g c t t	t c c c g c c c t t	a a t t c g g g c a	g t c a c a t g g a	t g g a t a a t c c	t a c a g a a g t a	360
t a t g t t a a t g	a t a g t g t a t g	g g t a c c t t g g c	c c c a t a g a t g	a t c g c t g c c c	t g c c a a a c c t	420
g a g g a a g a a g	g g a t g a t g a t	a a a t t t t c c	a t t g g g t a t c	a t t a t c t c c	t a t t g c c t a	480
g g g a g a g c a c	c a g g a t g t t t	a a t g c c t g c a	g t c c a a a a t t	g g t t g g t a g a	a g t a c t a c t	540
g t c a g t c c c a	t c t g t a g a t t	c a c t t a t c a c	a t g g t a a g c g	g g a t g t c a c t	c a g g c c a c g g	600
g t a a a t t a t t	t a c a a g a c t t	t t c t t a t c a a	a g a t c a t t a a	a a t t a g a c c	t a a a g g g a a a	660
c c t t g c c c c a	a g g a a t t t c c	c a a a g a t c a	a a a a a t a c a g	a a g t t t t a g t	t t g g g a a g a a	720
t g t g t g c c a	a t a g t g c g t	g a t t a t c a a	a a c a a t g a a t	t c g g a a c t a t	t a t a g a t t g g	780
g c a c c t c g a g	g t c a a t t c t a	c c a c a a t t g c	t c a g g a c a a a	c t c a g t c g t g	t c a a a g t g c a	840
c a a g t g a g t c	c a g c t g t t g a	t a g c a c t t a	a c a g a a a g t t	t a g a c a a a c a	t a a g c a t a a a	900
a a a t t g c a g t	c t t t c t a c c c	t t g g g a a t g g	g g a g a a a a a g	g a a t c t c a c	c c c a a g a c c a	960
a a a a t a g t a a	g t c c t g t t t c	t g g t c t g a a	c a t c c a g a a t	t a t g g a g g c t	t a c t g t g g c c	1020
t c a c a c c a c a	t t g a a t t t g	g t c t g g a a t	c a a a c t t t a g	a a a c a a g a g a	t c g t a a g c c a	1080
t t t t a t a c t a	t t g a c c t g a a	t t c c a g t c t a	a c a g t t c t t	t a c a a a g t t g	c g t a a a g g c c	1140
c c t t a t a t g c	t a g t t g t a g g	a a a t a t a g t t	a t t a a a c c a g	a c t c c c a g a c	t a t a a c t t g t	1200
g a a a a t t g t a	g a t t g c t t a c	t t g c a t t g a t	t c a a c t t t t a	a t t g g c a a c a	c c g t a t t c t g	1260
c t g g t g a g a g	c a a g a g a g g g	c g t g t g g a t c	c c t g t g t c c a	t g g a c c g a c c	g t g g g a g g c c	1320
t c g c c a t c c g	t c c a t a t t t a	g a c t g a a g t a	t t a a a g g t g	t t t t a a a t a g	a t c c a a a a g a	1380
t t c a t t t t a	c t t t a a t t g c	a g t g a t t g	g g a t t a a t t g	c a g t c a c a g c	t a c g g c t g c t	1440
g t a g c a g g	t t g c a t t g c a	c t c t t c t g t t	c a g t c a g t a a	a c t t g t t a a	t g a t t g g c a a	1500
a a a a a t t c t a	c a a g a t t g t g	g a a t t c a c a a	t c t a g t a t t g	a t c c a a a a t t	g g c a a a t c a a	1560
a t t a a t g a t c	t t a g a c a a a c	t g t c a t t t g g	a t g g g a g a c a	g a c t c a t g a g	c t t a g a c a t	1620
c g t t c a g t	t a c a a t g t g a	c t g c a t t t g g	t c a g a t t t t t	g t a t t a c c	c c a a a t t t a t	1680
a a t g a g t c t g	a g c a t c a c t g	g g a c a t g g t t	a g a c g c c a t c	t a c a g g g a a g	a g a a g a t a a t	1740
c t c a c t t a g	a c a t t t c c a a	a t t a a a a g a a	c a a a t t t c g	a a g c a t c a a a	a g c c c a t t a	1800
a a t t t g g t c	c a g g a c t g a	g g c a a t t g c a	g g a g t t g c t g	a t g c c t c g c	a a a t c t t a a c	1860
c c t g t c a c t t	g g g t t a a g a c	c a t t g g a a g t	a c t a c g a t t a	t a a t c t c a t	a t t a a t c t t	1920
g t g t g c c t g t	t t t g t c t g t t	g t t a g t c t g c	a g g t g t a c c	a a c a g c t c c g	a a g a g a c a g c	1980
g a c c a t c g a g	a a c g g g c c a t	g a t g a c g a t g	g c g g t t t t g t	c g a a a a g a a a	a g g g g g a a a t	2040
g t g g g a a a a a	g c a a g a g a g a	t c a g a t t g t t	a c t g t g t c t g	t g t a g		2085

<210> 24

<211> 1665

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 24

g t c a c a t g g a	t g g a t a a t c c	t a t g a a g t a	t a t g t t a a t g	a t a g t g t a t g	g g t a c c t t g g c	60
c c c a c a g a t g	a t c g c t g c c c	t g c c a a a c c t	g a g g a a g a a g	g g a t g a t g a t	a a a t t t t c c	120
a t t g t t a t c	g t t a t c c t c c	t a t t g c c t a	g g g a g a g c a c	c a g g a t g t t t	a a t g c c t g c a	180
g t c c a a a a t t	g g t t g g t a g a	a g t a c t a c t	g t c a g t c c t a	a c a g t a g a t t	c a c t t a t c a c	240
a t g g t a a g c g	g g a t g t c a c t	c a g g c c a c g g	g t a a a t t a t t	t a c a a g a c t t	t t c t t a t c a a	300

Substitute Sequence Listing_USSN 10587032_PP019482.007

agatcattaa	aat tagacc	taaaggaaaa	ccttgc cccca	aggaaaattcc	caaagaatca	360
aaaaatacag	agttttagt	ttgggaagaa	tgtgtggcca	atagtgcggt	gatattacaa	420
aacaatgaat	tcggaactat	tatagattgg	gcacctcgag	gtcaattcta	ccacaattgc	480
tcaggacaaa	ctcagtcgt	tccaa gtc	caagtga gtc	cagctgttga	tagcactta	540
acagaaa gtc	tagaca aaca	taagcataaa	aaattacagt	cttctaccc	ttgggaatgg	600
ggagaaaaag	gaatctctac	cccaagacca	gaaaataataa	gtcctgttgc	tggtccctgaa	660
catccagaat	tatggaggt	ttggcctgac	accacattag	aatttggtct	ggaaatcaaa	720
ctttagaaac	aagagatcg	aaggcatttt	atactatcg	cctaaattcc	agtcta acgg	780
ttccttaca	aagg tgcgt	aaggcctctt	atatgctagt	tgttaggaaat	atagttatta	840
aaccagactc	ccaaactata	acctgtaaa	attgttagatt	gttacttgc	attgattcaa	900
cttttaattg	gcccgc accgt	attctgctgg	tgagagcaag	agagggcgtg	tggatctctg	960
tgtccgtgga	ctgaccgtgg	gaggcctcgc	catccatcca	tat tttgact	gaagtattaa	1020
aagacatttt	aaatagatcc	aaaagattca	tttttaccc	aattgcagt	attatggat	1080
taattgcagt	cacagctacg	gctgctgtgg	caggagtgc	attgcactct	tctgttcagt	1140
cggtaaactt	tgttaatgat	tgccaaaaga	attctacaag	attgtggat	tcacaatcta	1200
gtattgatca	aaaattggca	aatcaaatta	atgatcttag	acaaactgtc	atttggatgg	1260
gagacagact	catgagctt	gaacattgtt	tccagttaca	gttgactgg	aatacgtcag	1320
at ttttgtat	taaccccaa	atttataatg	agtctgagca	tcactgggac	atggtagac	1380
gccatcata	ggaaagagaa	gataatctca	ctttagacat	ttccaaattt	aaataacaaa	1440
tttcaagc	atcaaaagcc	catttaattt	tgatgccagg	aactgaggca	attgcaggag	1500
ttgctgatgg	cctcgcaat	cttaaccctt	tcacttgggt	taagaccatc	ggaagtacta	1560
tgattataaa	tctcatat	atccctgtgt	gcctgtttt	tctgttgtt	gtctgcagg	1620
gtacccaaca	gctccgaaga	gacagc gacc	atc gagaacg	ggcca		1665

<210> 25

<211> 4086

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 25

atggggcctc	tccaa ccccg	gttgc ctct	ccggccatga	tccaaaaga	ttggc ttta	60
attataattg	atctaaagga	ttgctttttt	accatccctc	tggcagagca	ggatttgaa	120
aaatttgcct	ttactatacc	agccataaaat	aataaagaac	cagccaccag	gtttcagtgg	180
aaagtgttac	ctcaggaaat	gcttaatagt	ccaaactattt	gtcagacttt	tgtaggtcga	240
gctctcaac	cagtgagaga	aaagtttca	gactgttata	ttattcatta	tattgatgat	300
at tttatgtt	ctgcagaaac	gaaagataaa	ttaattgact	gttatacatt	tctgcaagca	360
gaggttgc	atgctggact	ggcaatagca	tccgataaga	tccaaacctc	tactcccttt	420
cattattttag	ggatgcagat	agaaaataga	aaaattaac	cacaaaaat	agaaataaga	480
aaagacacat	taaaaacact	aatgattttt	caaaaaat	taggagat	taattgatt	540
cggccaa ctc	taggcatttcc	tacttgc	atgtcaa	tgttcttat	cttaagagga	600
gactcagact	taaatagtca	aagaatatta	accccgagg	caacaaaaga	aattaaattt	660
gtggaaagaaa	aaattcagtc	agcgc aaata	aatagaatag	atcccttagc	cccactccaa	720
ctttt gattt	ttgccactgc	acattctca	acaggcatca	ttattcaaaa	tactgatctt	780
gtggagtgg	cattcccttcc	tcacagtaca	gttaagactt	ttacattgtt	cttggatcaa	840
atagctacat	taatcggtca	gacaagatta	cgaataacaa	aattatgtgg	aatgacccca	900
gacaaaatag	ttgtcccttt	aaccaaggaa	caagtttagc	aaggctttat	caattctggt	960
gcatggcaga	ttggcttgc	taattttgtt	ggacttattt	ataatcatta	cccaaaaaca	1020
aagatcttcc	agttcttaaa	attgactact	tggattctac	ctaaaattac	cagacgtgaa	1080
cctttagaaa	atgctctaa	agtattttact	gatggtttca	gcaatggaaa	agcagcttac	1140
acagggcccga	aagaacgagt	aatcaaaact	ccatatcaat	cggctcaag	agacgagttg	1200
gttgcagtca	ttacagtgtt	acaagatttt	gaccaaccta	tcaatattat	atcagattct	1260
gcataatgt	tacaggctac	aaggatgtt	gagacagctc	taattaaata	tagcatggat	1320
gatcagttaa	accagctatt	caatttatta	caacaaactg	taagaaaaag	aaatttccca	1380
ttttatatta	cttatatttc	agcacacact	aatttaccag	ggccttgc	taaagcaa at	1440
gaacaagctg	acttactgg	atcatctgc	ctcataaaag	cacaagaact	tcatgcttt	1500
actcatgtaa	atgcagcagg	attaaaaaac	aaatttgc	tcacatggaa	acaggcaaaa	1560
gatattgtac	aacattgcac	ccagtgtaa	gtcttacacc	tgcccactca	agaggcagga	1620
gttaatccca	gagggtctgt	tcctaatgc	ttatggcaaa	tggatgtc	gcatgtacct	1680
tcatttggaa	gattatcata	tgttcatgt	acagttgata	cttattcaca	tttcatatgg	1740
gcaacttgc	aaacaggaga	aagtacttcc	catgtaaaa	aacatttt	gtcttgg	1800
gctgtaatgg	gagttccaga	aaaaatcaaa	actgacaatg	gaccaggata	ttgttagaaa	1860
gctttccaaa	aattcttaag	tcagtgaaa	atttcacata	caacaggaat	tccttataat	1920
tcccaaggac	aggccatagt	tga aagaact	aatagaacac	tcaaaaactca	attagtaaa	1980

Substitute Sequence Listing_USSN 10587032_PP019482.007

caaaaagaag	ggggagacag	taaggagtgt	accactcctc	agatgcaact	taatctagca	2040
ctctatactt	taaattttt	aaacatttat	agaaatcaga	ctactacttc	tgcagaacaa	2100
catcttactg	gtaaaaaagaa	cagcccacat	gaaggaaaac	taatttggtg	gaaagataat	2160
aaaaataaga	catgggaaat	agggaaagggt	ataacgtggg	ggagaggttt	tgcttggtt	2220
tcaccaggag	aaaatcagct	tcctgtttgg	ttacccacta	gacatttcaa	gttctacaat	2280
gaaccatcg	gagatgcaaa	gaaaagggcc	tccacggaga	tggtaacacc	agtcacatgg	2340
atggataatc	ctatagaagt	atatgttaat	gatagtatat	gggtacctgg	ccccatagat	2400
gatcgctgcc	ctgccaaacc	tgaggaagaa	gggatgtga	taaatatttc	cattgggtat	2460
cgttatcctc	ctattingcct	agggagagca	ccaggatgtt	taatgcctgc	agtccaaaat	2520
tggtttgttag	aagtacctac	tgtcagtccc	atcagtagat	tcacttatca	catggtaagc	2580
gggatgtcac	tcaggccacg	ggtaaattat	ttacaagact	tttcttatca	aagatcatta	2640
aaatttagac	ctaaagggaa	accttgcccc	aaggaaattc	ccaaagaatc	aaaaaataca	2700
gaagttttag	tttgggaaaga	atgtgtggcc	aatagtgcgg	tgatattata	aaacaatgaa	2760
tttggaaacta	ttatagattg	ggcacctcga	ggtaattct	accacaattt	ctcaggacaa	2820
actcagtcgt	gtccaagtgc	acaagtgggt	ccagctttt	atagcgactt	aacagaaagt	2880
ttagacaaac	ataagcataa	aaaattgcag	tcttctacc	cttgggaatg	gggagaaaaaa	2940
ggaatctcta	ccccaaagacc	aaaatagta	agtccctgtt	ctggcctgt	acatccagaa	3000
ttatggagggc	ttactgtgc	ctcacaccac	attagaattt	ggtttggaaa	tcaacttta	3060
gaaacaagag	attgtaaagcc	atttatact	gtcgacctaa	attccagtct	aacagtccct	3120
ttacaaggt	gcgttaaaggc	cccttataatg	ctagttttag	gaatataatgt	tattaaacca	3180
gactcccaga	ctataacctg	tgaaaattgt	agattgcctt	cttgcattga	ttcaacttta	3240
aattggcaac	accgtattct	gctggtgaga	gcaagagagg	gcgtgtggat	ccctgtgtcc	3300
atggaccgac	cgtgggaggg	ctcaccatcc	gtccatattt	tgactgaagt	attaaaaggt	3360
gttttaataa	gatccaaaag	attcattttt	actttaattt	cagtgattat	gggattaatt	3420
gcagtcacag	ctacggctgc	tgttagcagga	gttgcattgc	actcttctgt	tcagtcaat	3480
aactttgtta	atgattggca	aaagaattct	acaagattgt	ggaattcaca	atctgtatt	3540
gatcaaaaat	tggcaaatca	aattaatgtat	cttagacaaa	ctgtcattt	gatgggagac	3600
agactcatga	gcttagaaca	tcgttccag	ttacaatgt	actggaataac	gtcagatttt	3660
tgtattacac	cccaaattta	taatgagtct	gagcatcaat	gggacatgg	tagacccat	3720
ctacaggaa	gagaagataa	tctcaactta	gacatttcca	aattaaaaga	acaaattttc	3780
gaagcatcaa	aagcccattt	aaatttggtg	ccaggaactt	aggcaatttgc	aggagttgt	3840
gatggcctcg	caaatcttaa	ccctgtcact	tgggttaaga	ccattggaaag	tacatcgatt	3900
ataaatctca	tattaatctt	tgtgtgcctg	ttttgtctgt	tgttagtctg	caggtgtacc	3960
caacagctcc	gaagagagacag	cgaccatcga	gaacggggca	tgatgacgat	ggcggtttt	4020
tcgaaaagaa	aagggggaaa	tgtggggaaa	agcaagagag	atcaaattgt	tactgtgtct	4080
						4086

<210> 26

<211> 694

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 26

Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg His Arg Asn Arg Ala
1 5 10 15

Pro Leu Thr His Lys Met Asn Lys Met Val Thr Ser Glu Glu Gln Met
20 25 30

Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro Thr Trp Ala Gln Leu
35 40 45

Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu Glu Asn Thr Lys Val
50 55 60

Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala Leu Met Ile Val Ser
65 70 75 80

Met Val Val Ser Leu Pro Met Pro Ala Gly Ala Ala Ala Asn Tyr
85 90 95

Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu Ile Arg Ala Val Thr
100 105 110

Substitute Sequence Listing_USSN 10587032_PP019482.007

Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn Asp Ser Val Trp Val
115 120 125

Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys Pro Glu Glu Glu Gly
130 135 140

Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr Pro Pro Ile Cys Leu
145 150 155 160

Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val Gln Asn Trp Leu Val
165 170 175

Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe Thr Tyr His Met Val
180 185 190

Ser Gly Met Ser Leu Arg Pro Arg Val Asn Tyr Leu Gln Asp Phe Ser
195 200 205

Tyr Gln Arg Ser Leu Lys Phe Arg Pro Lys Gly Lys Pro Cys Pro Lys
210 215 220

Glu Ile Pro Lys Glu Ser Lys Asn Thr Glu Val Leu Val Trp Glu Glu
225 230 235 240

Cys Val Ala Asn Ser Ala Val Ile Leu Gln Asn Asn Glu Phe Gly Thr
245 250 255

Ile Ile Asp Trp Ala Pro Arg Gly Gln Phe Tyr His Asn Cys Ser Gly
260 265 270

Gln Thr Gln Ser Cys Gln Ser Ala Gln Val Ser Pro Ala Val Asp Ser
275 280 285

Asp Leu Thr Glu Ser Leu Asp Lys His Lys His Lys Leu Gln Ser
290 295 300

Phe Tyr Pro Trp Glu Trp Gly Glu Lys Gly Ile Ser Thr Pro Arg Pro
305 310 315 320

Lys Ile Val Ser Pro Val Ser Gly Pro Glu His Pro Glu Leu Trp Arg
325 330 335

Leu Thr Val Ala Ser His His Ile Arg Ile Trp Ser Gly Asn Gln Thr
340 345 350

Leu Glu Thr Arg Asp Arg Lys Pro Phe Tyr Thr Ile Asp Leu Asn Ser
355 360 365

Ser Leu Thr Val Pro Leu Gln Ser Cys Val Lys Pro Pro Tyr Met Leu
370 375 380

Val Val Gly Asn Ile Val Ile Lys Pro Asp Ser Gln Thr Ile Thr Cys
385 390 395 400

Glu Asn Cys Arg Leu Leu Thr Cys Ile Asp Ser Thr Phe Asn Trp Gln
405 410 415

His Arg Ile Leu Leu Val Arg Ala Arg Glu Gly Val Trp Ile Pro Val
420 425 430

Ser Met Asp Arg Pro Trp Glu Ala Ser Pro Ser Val His Ile Leu Thr
435 440 445

Substitute Sequence Listing_USSN 10587032_PP019482.007

Glu Val Leu Lys Gly Val Leu Asn Arg Ser Lys Arg Phe Ile Phe Thr
450 455 460
Leu Ile Ala Val Ile Met Gly Leu Ile Ala Val Thr Ala Thr Ala Ala
465 470 475 480
Val Ala Gly Val Ala Leu His Ser Ser Val Gln Ser Val Asn Phe Val
485 490 495
Asn Asp Trp Gln Lys Asn Ser Thr Arg Leu Trp Asn Ser Gln Ser Ser
500 505 510
Ile Asp Gln Lys Leu Ala Asn Gln Ile Asn Asp Leu Arg Gln Thr Val
515 520 525
Ile Trp Met Gly Asp Arg Leu Met Ser Leu Glu His Arg Phe Gln Leu
530 535 540
Gln Cys Asp Trp Asn Thr Ser Asp Phe Cys Ile Thr Pro Gln Ile Tyr
545 550 555 560
Asn Glu Ser Glu His His Trp Asp Met Val Arg Arg His Leu Gln Gly
565 570 575
Arg Glu Asp Asn Leu Thr Leu Asp Ile Ser Lys Leu Lys Glu Gln Ile
580 585 590
Phe Glu Ala Ser Lys Ala His Leu Asn Leu Val Pro Gly Thr Glu Ala
595 600 605
Ile Ala Gly Val Ala Asp Gly Leu Ala Asn Leu Asn Pro Val Thr Trp
610 615 620
Val Lys Thr Ile Gly Ser Thr Thr Ile Ile Asn Leu Ile Leu Ile Leu
625 630 635 640
Val Cys Leu Phe Cys Leu Leu Leu Val Cys Arg Cys Thr Gln Gln Leu
645 650 655
Arg Arg Asp Ser Asp His Arg Glu Arg Ala Met Met Thr Met Ala Val
660 665 670
Leu Ser Lys Arg Lys Gly Gly Asn Val Gly Lys Ser Lys Arg Asp Gln
675 680 685
Ile Val Thr Val Ser Val
690

<210> 27
<211> 1361
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<220>
<221> SITE
<222> 917
<223> Xaa is any amino acid

<400> 27
Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met Ile Pro Lys
1 5 10 15

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asp Trp Pro Leu Ile Ile Ile Asp Leu Lys Asp Cys Phe Phe Thr Ile
20 25 30

Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr Ile Pro Ala
35 40 45

Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys Val Leu Pro
50 55 60

Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe Val Gly Arg
65 70 75 80

Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr Ile Ile His
85 90 95

Tyr Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp Lys Leu Ile
100 105 110

Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala Gly Leu Ala
115 120 125

Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His Tyr Leu Gly
130 135 140

Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile Glu Ile Arg
145 150 155 160

Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu Leu Gly Asp
165 170 175

Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr Ala Met Ser
180 185 190

Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn Ser Gln Arg
195 200 205

Ile Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val Glu Glu Lys
210 215 220

Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala Pro Leu Gln
225 230 235 240

Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile Ile Ile Gln
245 250 255

Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser Thr Val Lys
260 265 270

Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile Gly Gln Thr
275 280 285

Arg Leu Arg Ile Thr Lys Leu Cys Gly Asn Asp Pro Asp Lys Ile Val
290 295 300

Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile Asn Ser Gly
305 310 315 320

Ala Trp Gln Ile Gly Leu Ala Asn Phe Val Gly Leu Ile Asp Asn His
325 330 335

Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr Thr Trp Ile
340 345 350

Substitute Sequence Listing_USSN 10587032_PP019482.007

Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala Leu Thr Val
355 360 365

Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr Gly Pro Lys
370 375 380

Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg Asp Glu Leu
385 390 395 400

Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro Ile Asn Ile
405 410 415

Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp Val Glu Thr
420 425 430

Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln Leu Phe Asn
435 440 445

Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe Tyr Ile Thr
450 455 460

Tyr Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Lys Ala Asn
465 470 475 480

Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys Ala Gln Glu
485 490 495

Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys Asn Lys Phe
500 505 510

Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His Cys Thr Gln
515 520 525

Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val Asn Pro Arg
530 535 540

Gly Leu Cys Pro Asn Ala Leu Trp Gln Met Asp Val Thr His Val Pro
545 550 555 560

Ser Phe Gly Arg Leu Ser Tyr Val His Val Thr Val Asp Thr Tyr Ser
565 570 575

His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser His Val
580 585 590

Lys Lys His Leu Leu Ser Cys Phe Ala Val Met Gly Val Pro Glu Lys
595 600 605

Ile Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Phe Gln Lys
610 615 620

Phe Leu Ser Gln Trp Lys Ile Ser His Thr Thr Gly Ile Pro Tyr Asn
625 630 635 640

Ser Gln Gly Gln Ala Ile Val Glu Arg Thr Asn Arg Thr Leu Lys Thr
645 650 655

Gln Leu Val Lys Gln Lys Glu Gly Gly Asp Ser Lys Glu Cys Thr Thr
660 665 670

Pro Gln Met Gln Leu Asn Leu Ala Leu Tyr Thr Leu Asn Phe Leu Asn
675 680 685

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ile Tyr Arg Asn Gln Thr Thr Ser Ala Glu Gln His Leu Thr Gly
690 695 700

Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp Asn
705 710 715 720

Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Thr Trp Gly Arg Gly
725 730 735

Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Leu Pro
740 745 750

Thr Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Gly Asp Ala Lys Lys
755 760 765

Arg Ala Ser Thr Glu Met Val Thr Pro Val Thr Trp Met Asp Asn Pro
770 775 780

Ile Glu Val Tyr Val Asn Asp Ser Ile Trp Val Pro Gly Pro Ile Asp
785 790 795 800

Asp Arg Cys Pro Ala Lys Pro Glu Glu Gly Met Met Ile Asn Ile
805 810 815

Ser Ile Gly Tyr Arg Tyr Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly
820 825 830

Cys Leu Met Pro Ala Val Gln Asn Trp Leu Val Glu Val Pro Thr Val
835 840 845

Ser Pro Ile Ser Arg Phe Thr Tyr His Met Val Ser Gly Met Ser Leu
850 855 860

Arg Pro Arg Val Asn Tyr Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu
865 870 875 880

Lys Phe Arg Pro Lys Gly Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu
885 890 895

Ser Lys Asn Thr Glu Val Leu Val Trp Glu Glu Cys Val Ala Asn Ser
900 905 910

Ala Val Ile Leu Xaa Asn Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala
915 920 925

Pro Arg Gly Gln Phe Tyr His Asn Cys Ser Gly Gln Thr Gln Ser Cys
930 935 940

Pro Ser Ala Gln Val Ser Pro Ala Val Asp Ser Asp Leu Thr Glu Ser
945 950 955 960

Leu Asp Lys His Lys His Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu
965 970 975

Trp Gly Glu Lys Gly Ile Ser Thr Pro Arg Pro Lys Ile Val Ser Pro
980 985 990

Val Ser Gly Pro Glu His Pro Glu Leu Trp Arg Leu Thr Val Ala Ser
995 1000 1005

His His Ile Arg Ile Trp Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp
1010 1015 1020

Substitute Sequence Listing_USSN 10587032_PP019482.007

Cys Lys Pro Phe Tyr Thr Val Asp Leu Asn Ser Ser Leu Thr Val Pro
1025 1030 1035 1040

Leu Gln Ser Cys Val Lys Pro Pro Tyr Met Leu Val Val Gly Asn Ile
1045 1050 1055

Val Ile Lys Pro Asp Ser Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu
1060 1065 1070

Leu Thr Cys Ile Asp Ser Thr Phe Asn Trp Gln His Arg Ile Leu Leu
1075 1080 1085

Val Arg Ala Arg Glu Gly Val Trp Ile Pro Val Ser Met Asp Arg Pro
1090 1095 1100

Trp Glu Ala Ser Pro Ser Val His Ile Leu Thr Glu Val Leu Lys Gly
1105 1110 1115 1120

Val Leu Asn Arg Ser Lys Arg Phe Ile Phe Thr Leu Ile Ala Val Ile
1125 1130 1135

Met Gly Leu Ile Ala Val Thr Ala Ala Val Ala Gly Val Ala
1140 1145 1150

Leu His Ser Ser Val Gln Ser Val Asn Phe Val Asn Asp Trp Gln Lys
1155 1160 1165

Asn Ser Thr Arg Leu Trp Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu
1170 1175 1180

Ala Asn Gln Ile Asn Asp Leu Arg Gln Thr Val Ile Trp Met Gly Asp
1185 1190 1195 1200

Arg Leu Met Ser Leu Glu His Arg Phe Gln Leu Gln Cys Asp Trp Asn
1205 1210 1215

Thr Ser Asp Phe Cys Ile Thr Pro Gln Ile Tyr Asn Glu Ser Glu His
1220 1225 1230

His Trp Asp Met Val Arg Arg His Leu Gln Gly Arg Glu Asp Asn Leu
1235 1240 1245

Thr Leu Asp Ile Ser Lys Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys
1250 1255 1260

Ala His Leu Asn Leu Val Pro Gly Thr Glu Ala Ile Ala Gly Val Ala
1265 1270 1275 1280

Asp Gly Leu Ala Asn Leu Asn Pro Val Thr Trp Val Lys Thr Ile Gly
1285 1290 1295

Ser Thr Ser Ile Ile Asn Leu Ile Leu Ile Leu Val Cys Leu Phe Cys
1300 1305 1310

Leu Leu Leu Val Cys Arg Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp
1315 1320 1325

His Arg Glu Arg Ala Met Met Thr Met Ala Val Leu Ser Lys Arg Lys
1330 1335 1340

Gly Gly Asn Val Gly Lys Ser Lys Arg Asp Gln Ile Val Thr Val Ser
1345 1350 1355 1360

Val

Substitute Sequence Listing_USSN 10587032_PP019482.007

<210> 28
<211> 699
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 28
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65 70 75 80

Leu Met Ile Val Ser Met Val Val Ser Leu Pro Met Pro Ala Gly Ala
85 90 95

Ala Ala Ala Asn Tyr Thr Trp Ala Tyr Val Pro Phe Pro Pro Leu
100 105 110

Ile Arg Ala Val Thr Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn
115 120 125

Asp Ser Val Trp Val Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys
130 135 140

Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr
145 150 155 160

Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val
165 170 175

Gln Asn Trp Leu Val Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe
180 185 190

Thr Tyr His Met Val Ser Gly Met Ser Leu Arg Pro Arg Val Asn Tyr
195 200 205

Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu Lys Phe Arg Pro Lys Gly
210 215 220

Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu Ser Lys Asn Thr Glu Val
225 230 235 240

Leu Val Trp Glu Glu Cys Val Ala Asn Ser Ala Val Ile Leu Gln Asn
245 250 255

Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala Pro Arg Gly Gln Phe Tyr
260 265 270

His Asn Cys Ser Gly Gln Thr Gln Ser Cys Pro Ser Ala Gln Val Ser
275 280 285

Substitute Sequence Listing_USSN 10587032_PP019482.007

Pro Ala Val Asp Ser Asp Leu Thr Glu Ser Leu Asp Lys His Lys His
290 295 300

Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu Trp Gly Glu Lys Gly Ile
305 310 315 320

Ser Thr Pro Arg Pro Lys Ile val Ser Pro val Ser Gly Pro Glu His
325 330 335

Pro Glu Leu Trp Arg Leu Thr Val Ala Ser His His Ile Arg Ile Trp
340 345 350

Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp Arg Lys Pro Phe Tyr Thr
355 360 365

Ile Asp Leu Asn Ser Ser Leu Thr Val Pro Leu Gln Ser Cys Val Lys
370 375 380

Pro Pro Tyr Met Leu Val Val Gly Asn Ile Val Ile Lys Pro Asp Ser
385 390 395 400

Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu Leu Thr Cys Ile Asp Ser
405 410 415

Thr Phe Asn Trp Gln His Arg Ile Leu Leu Val Arg Ala Arg Glu Gly
420 425 430

Val Trp Ile Pro Val Ser Met Asp Arg Pro Trp Glu Ala Ser Pro Ser
435 440 445

Val His Ile Leu Thr Glu Val Leu Lys Gly Val Leu Asn Arg Ser Lys
450 455 460

Arg Phe Ile Phe Thr Leu Ile Ala Val Ile Met Gly Leu Ile Ala Val
465 470 475 480

Thr Ala Thr Ala Ala Val Ala Gly Val Ala Leu His Ser Ser Val Gln
485 490 495

Ser Val Asn Phe Val Asn Asp Trp Gln Lys Asn Ser Thr Arg Leu Trp
500 505 510

Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu Ala Asn Gln Ile Asn Asp
515 520 525

Leu Arg Gln Thr Val Ile Trp Met Gly Asp Arg Leu Met Ser Leu Glu
530 535 540

His Arg Phe Gln Leu Gln Cys Asp Trp Asn Thr Ser Asp Phe Cys Ile
545 550 555 560

Thr Pro Gln Ile Tyr Asn Glu Ser Glu His His Trp Asp Met Val Arg
565 570 575

Arg His Leu Gln Gly Arg Glu Asp Asn Leu Thr Leu Asp Ile Ser Lys
580 585 590

Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys Ala His Leu Asn Leu Val
595 600 605

Pro Gly Thr Glu Ala Ile Ala Gly Val Ala Asp Gly Leu Ala Asn Leu
610 615 620

Substitute Sequence Listing_USSN 10587032_PP019482.007
Asn Pro Val Thr Trp Val Lys Thr Ile Gly Ser Thr Thr Ile Ile Asn
625 630 635 640

Leu Ile Leu Ile Leu Val Cys Leu Phe Cys Leu Leu Leu Val Cys Arg
645 650 655

Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp His Arg Glu Arg Ala Met
660 665 670

Met Thr Met Ala Val Leu Ser Lys Arg Lys Gly Gly Asn Val Gly Lys
675 680 685

Ser Lys Arg Asp Gln Ile Val Thr Val Ser Val
690 695

<210> 29
<211> 294

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 29

agttctacaa tgaacccatc agagatgcaa agaaaagcac ctccgcggag acggagacat 60
cgcaatcgag caccgttgac tcacaagatg aacaaaatgg tgacgtcaga agaacagatg 120
aagttgccat ccaccaagaa ggcagagccg ccaacttggg cacaactaaa gaagctgacg 180
cagttagcta caaaatatct agagaacaca aaggtgacac aaaccccaga gagtatgctg 240
cttgcagcct ttagtattgt atcaatggtg gtaagtctcc ctatgcctgc agga 294

<210> 30

<211> 57

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 30

tctgcaggta tacccaacag ctccgaagag acagcgacca tcgagaacgg gccatga 57

<210> 31

<211> 105

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 31

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 . 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Gly Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65 70 75 80

Leu Met Ile Val Ser Met Val Ser Ala Gly Val Pro Asn Ser Ser Glu
85 90 95

Glu Thr Ala Thr Ile Glu Asn Gly Pro
100 105

<210> 32

Substitute Sequence Listing_USSN 10587032_PP019482.007

<211> 86

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 32

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Ser Ala Gly Val Pro Asn Ser Ser Glu Glu Thr Ala
65 70 75 80

Thr Ile Glu Asn Gly Pro
85

<210> 33

<211> 74

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 33

Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
1 5 10 15

Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
20 25 30

His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
35 40 45

Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
50 55 60

Arg His Arg Arg Leu His Phe Val Leu Tyr
65 70

<210> 34

<211> 79

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 34

Met Asn Ser Leu Glu Met Gln Arg Lys Val Trp Arg Trp Arg His Pro
1 5 10 15

Asn Arg Leu Ala Ser Leu Gln Val Tyr Pro Ala Ala Pro Lys Arg Gln
20 25 30

Gln Pro Ala Arg Met Gly His Ser Asp Asp Gly Gly Phe Val Lys Lys
35 40 45

Lys Arg Gly Gly Tyr Val Arg Lys Arg Glu Ile Arg Leu Ser Leu Cys
50 55 60

Leu Cys Arg Lys Gly Arg His Lys Lys Leu His Phe Val Leu Tyr

65

Substitute Sequence Listing_USSN 10587032_PP019482.007
70 75<210> 35
<211> 129

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 35

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60Glu Asn Thr Lys Val Ile Leu Gln Val Tyr Pro Thr Ala Pro Lys Arg
65 70 75 80Gln Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu
85 90 95Lys Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr
100 105 110Cys Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu
115 120 125

Tyr

<210> 36
<211> 125

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 36

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60Glu Asn Thr Lys Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser
65 70 75 80Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly
85 90 95Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val
100 105 110

Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr

Substitute Sequence Listing_USSN 10587032_PP019482.007
115 120 125

<210> 37
<211> 144
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 37
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65 70 75 80

Leu Met Ile Val Ser Met Val Val Tyr Pro Thr Ala Pro Lys Arg Gln
85 90 95

Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys
100 105 110

Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys
115 120 125

Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr
130 135 140

<210> 38
<211> 74
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 38
Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
1 5 10 15

Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
20 25 30

His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
35 40 45

Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
50 55 60

Arg His Arg Arg Leu His Phe Val Leu Tyr
65 70

<210> 39
<211> 74
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 39

Substitute Sequence Listing_USSN 10587032_PP019482.007
Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
1 5 10 15

Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
20 25 30

His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
35 40 45

Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
50 55 60

Arg His Arg Arg Leu His Phe Val Leu Tyr
65 70

<210> 40

<211> 44

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 40

Met Glu Tyr Lys Asn Arg His Leu Lys Phe Tyr Asn Glu Pro Ile Gly
1 5 10 15

Asp Ala Lys Lys Arg Ala Ser Thr Glu Met Ser Ala Gly Val Pro Asn
20 25 30

Ser Ser Glu Glu Thr Ala Thr Ile Glu Asn Gly Pro
35 40

<210> 41

<211> 74

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 41

Met Asn Pro Ser Glu Met Gln Arg Lys Gly Pro Pro Gln Arg Cys Leu
1 5 10 15

Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser Arg Thr Gly
20 25 30

His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly Lys Cys Gly
35 40 45

Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val Glu Arg Ser
50 55 60

Arg His Arg Arg Leu His Phe Val Leu Tyr
65 70

<210> 42

<211> 86

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 42

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Substitute Sequence Listing_USSN 10587032_PP019482.007
Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Ser Ala Gly Val Pro Asn Ser Ser Glu Glu Thr Ala
65 70 75 80

Thr Ile Glu Asn Gly Pro
85

<210> 43

<211> 105

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 43

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65 70 75 80

Leu Met Ile Val Ser Met Val Ser Ala Gly Val Pro Asn Ser Ser Glu
85 90 95

Glu Thr Ala Thr Ile Glu Asn Gly Pro
100 105

<210> 44

<211> 127

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 44

Met Val Thr Pro Val Thr Trp Met Asp Asn Pro Ile Glu Val Tyr Val
1 5 10 15

Asn Asp Ser Val Trp Val Pro Gly Pro Thr Asp Asp Arg Cys Pro Ala
20 25 30

Lys Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Val Tyr Arg
35 40 45

Tyr Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala
50 55 60

Val Gln Asn Cys Leu Gln Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg
65 70 75 80

Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys
85 90 95

Substitute Sequence Listing_USSN 10587032_PP019482.007

Arg	Gly	Lys	Cys	Gly	Glu	Lys	Gln	Glu	Arg	Ser	Asp	Cys	Tyr	Cys	Val
			100			105						110			

Cys	Val	Glu	Arg	Ser	Arg	His	Arg	Arg	Leu	His	Phe	Val	Leu	Tyr	
		115			120						125				

<210> 45
<211> 105
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400>	45														
Met	Val	Thr	Pro	Val	Thr	Trp	Met	Asp	Asn	Pro	Ile	Glu	Val	Tyr	Val
1							5		10				15		

Asn	Asp	Ser	Glu	Trp	Val	Pro	Gly	Pro	Thr	Asp	Asp	Arg	Cys	Pro	Ala
			20					25				30			

Lys	Pro	Glu	Glu	Glu	Gly	Met	Met	Ile	Asn	Ile	Ser	Ile	Gly	Leu	Gln
			35				40					45			

Val	Tyr	Pro	Thr	Ala	Pro	Lys	Arg	Gln	Arg	Pro	Ser	Arg	Thr	Gly	His
			50				55				60				

Asp	Asp	Asp	Gly	Gly	Phe	Val	Glu	Lys	Lys	Arg	Gly	Lys	Cys	Gly	Glu
			65				70		75			80			

Lys	Gln	Glu	Arg	Ser	Asp	Cys	Tyr	Cys	Val	Cys	Val	Glu	Arg	Ser	Arg
			85					90				95			

His	Arg	Arg	Leu	His	Phe	Val	Met	Cys							
			100				105								

<210> 46
<211> 79
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400>	46														
Met	Asn	Ser	Leu	Glu	Met	Gln	Arg	Lys	Val	Trp	Arg	Trp	Arg	His	Pro
1					5				10				15		

Asn	Arg	Leu	Ala	Ser	Leu	Gln	Val	Tyr	Pro	Ala	Ala	Pro	Lys	Arg	Gln
			20				25					30			

Gln	Pro	Ala	Arg	Met	Gly	His	Ser	Asp	Asp	Gly	Gly	Phe	Val	Lys	Lys
			35				40				45				

Lys	Arg	Gly	Gly	Tyr	Val	Arg	Lys	Arg	Glu	Ile	Arg	Leu	Ser	Leu	Cys
			50				55			60					

Leu	Cys	Arg	Lys	Gly	Arg	His	Lys	Lys	Leu	His	Phe	Asp	Leu	Tyr	
			65				70			75					

<210> 47
<211> 214
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400>	47														
Met	Asn	Ser	Leu	Glu	Met	Gln	Arg	Lys	Ala	Pro	Pro	Arg	Arg	Arg	
1					5				10			15			

Substitute Sequence Listing_USSN 10587032_PP019482.007
His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Ser Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65 70 75 80

Leu Met Ile Val Ser Met Val Val Ser Leu Pro Met Pro Ala Gly Ala
85 90 95

Ala Ala Ala Asn Tyr Thr Tyr Trp Ala Tyr Val Pro Phe Pro Pro Leu
100 105 110

Ile Arg Ala Val Thr Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn
115 120 125

Asp Ser Val Trp Val Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys
130 135 140

Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr
145 150 155 160

Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val
165 170 175

Gln Asn Trp Leu Val Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe
180 185 190

Thr Tyr His Met Ser Ala Gly Val Pro Asn Ser Ser Glu Glu Thr Ala
195 200 205

Thr Ile Glu Asn Gly Pro
210

<210> 48

<211> 129

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 48

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Leu Gln Val Tyr Pro Thr Ala Pro Lys Arg
65 70 75 80

Gln Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu
85 90 95

Substitute Sequence Listing_USSN 10587032_PP019482.007
Lys Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr
100 105 110

Cys Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Met
115 120 125

Tyr

<210> 49
<211> 125

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 49

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Tyr Pro Thr Ala Pro Lys Arg Gln Arg Pro Ser
65 70 75 80

Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu Lys Lys Arg Gly
85 90 95

Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys Val Cys Val
100 105 110

Glu Arg Ser Arg His Arg Arg Leu His Phe Val Met Tyr
115 120 125

<210> 50
<211> 145

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<220>

<221> SITE

<222> 64

<223> Xaa is any amino acid

<400> 50

Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Xaa
50 55 60

Leu Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala

Substitute Sequence Listing_USSN 10587032_PP019482.007
70 75 80

65

70

75

80

Ala Leu Met Ile Val Ser Met Val Val Tyr Pro Thr Ala Pro Lys Arg
85 90 95

Gln Arg Pro Ser Arg Thr Gly His Asp Asp Asp Gly Gly Phe Val Glu
100 105 110

Lys Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr
115 120 125

Cys Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Met
130 135 140

Tyr
145

<210> 51
<211> 4657
<212> DNA
<213> Artificial Sequence

<220>
<223> pCMVKm2.cORFopt HML-2 vector

<400>	51	60				
gcccggaaat	ttcgactcta	ggccattgca	tacgttgtat	ctatatcata	atatgtacat	60
tttatattggc	tcatgtccaa	tatgaccgccc	atgttgcacat	tgattattga	ctagttatta	120
atagtaatca	attacggggt	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acctacggta	aatggcccg	ctggctgacc	gcccaacgac	cccccccat	tgacgtcaat	240
aatgacgtat	gttcccata	taacgccaat	agggacttgc	cattgacgtc	aatgggttga	300
gtatTTacgg	taaactgccc	acttggcagt	acatcaagtg	tatcatatgc	caagtccccc	360
ccctattgac	gtcaatgacg	gtaaatggcc	cgcctggcat	tatgcccagt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtatttagtc	atcgctatta	ccatgggtat	480
gcgggtttgg	cagtacacca	atggggcgtgg	atagcggttt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtc	atggggat	gttttggcac	caaaatcaac	gggacttccc	600
aaaatgtcgt	aataaccccg	ccccgttgc	gcaaatgggc	gtttaggcgt	tacgggttga	660
ggtctatata	agcagagctc	gtttagtcaa	ccgtcagatc	gccttggagac	gccccccacg	720
ctgttttgc	ctccatagaa	gacaccggga	ccgatccagc	ctcccgcc	gggaacgcgt	780
cattggAACG	cggattcccc	gtgccaagag	tgacgtaaat	accgcctata	gactctatag	840
gcacacccct	ttggctctta	tgcatgttat	actgttttg	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataggt	gatggatag	cttagcctat	aggtgtgggt	tattgaccat	960
tattgaccac	tcccctattg	gtgacgatac	tttccattac	taatccataa	catggctt	1020
tgccacaact	atctctattg	gctatatgccc	aataactctgt	ccttcagaga	ctgacacgg	1080
ctctgtatTT	ttacaggat	gggtccatt	tattatttac	aaattcacat	atacaacaac	1140
gccgtcccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gccaatctcg	1200
ggtacgtgtt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
ttgtcccatg	cctccagcgg	ctcatggtcg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagtg	tgccgcacaa	ggccgtggcg	1380
gttagggat	tgttgcaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcagatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagcttag	ttgttgtatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acgggtggagg	gcagtgtagt	ctgagcagta	1560
ctctgtgtc	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttccct	1620
tccatggtc	ttttctgcag	taccgtcgt	cgacgcacc	atgaacccca	gcgagatgca	1680
gcccggcc	ccccccccc	gccggccca	ccgcacccgc	gccccctga	ccccacaagat	1740
gaacaagat	gtgaccagcg	aggagcagat	gaagctggcc	agcacaaga	aggccgagcc	1800
ccccacctgg	gcccgactga	agaagctgac	ccagctggcc	accaagtacc	tgggaaacac	1860
caaggtgacc	cagacccccc	agagcatgct	gctggccggc	ctgatgatcg	tgagcatggt	1920
gagcggccggc	gtgccccaa	gcagcgagga	gaccggccacc	atcgagaacg	gccccgctta	1980
agaatttcg	actcgagcaa	gtctagaaag	ccatggatat	cgatccact	acgcgttaga	2040
gctcgctgt	cagcctcgac	tgtgccttct	agttgccagc	catctgttgt	ttgccccctcc	2100
cccgtgcct	cattgaccct	ggaagggtgcc	actcccactg	tcctttccta	ataaaatgag	2160
gaaattgcat	cgcattgtct	gagtaggtgt	cattctattc	tgggggggtgg	ggtgggggcag	2220

Substitute Sequence Listing_USSN 10587032_PP019482.007

gacagcaagg	gggaggattg	ggaagacaat	agcagggggg	tggcgaaga	actccagcat	2280
gagatccccg	cgctggagga	tcatccagcc	ggcgtcccg	aaaacgattc	cgaagccaa	2340
cctttcatag	aaggcggcgg	tggaatcgaa	atctcgat	ggcaggttgg	gcgtcgctt	2400
gtcggtcatt	tcgaacccca	gagtcccgt	cagaagaact	cgtcaagaag	gcgatagaag	2460
gcatgcgcgt	gcaatcggg	agcggcgata	ccgtaaagca	cgaggaagcg	gtcagccat	2520
tcgcccggaa	gctttcagc	aatatcacgg	gtagccaacg	ctatgtcctg	atagcgtcc	2580
gccacaccca	gccggccaca	gtcgatgaat	ccagaaaagc	ggccattttc	caccatgata	2640
ttcggcaagc	aggcatcgcc	atgggtcagc	acgagatct	cgccgtcggg	catgcgcgc	2700
ttgagcctgg	cgaacagttc	ggctggcgcg	agcccctgtat	gctttcgtc	cagatcatcc	2760
tgatcgacaa	gaccggcttc	catccgagta	cgtgctcgct	cgatgcgtat	tttcgcttgg	2820
tggtcgaatg	ggcaggtagc	cggatcaagc	gtatgcagcc	gcccatttc	atcagccatg	2880
atggatactt	tctcggcagg	agcaagggtg	gatgacagga	gatcctgccc	cggcacttcg	2940
cccaatagca	gccagttcct	ccccgcttca	gtgacaacgt	cgagcacagc	tgcgcaagga	3000
acgcccgtcg	tggccagcca	cgatagccgc	gctgctcggt	cctgcagttc	attcagggca	3060
ccggacaggt	cggcttgcac	aaaaagaacc	gggcgcctt	gcgtgcacag	ccggaacacg	3120
gccccatcg	agcagccgt	tgtctgttgc	gcccagtcat	agccgaatag	cctctccacc	3180
caagcggccg	gagaacctgc	gtgcaatcca	tcttgttcaa	tcatgcgaaa	cgatccat	3240
cctgtcttctt	gatcagatct	tgatccccgt	cgccatcaga	tccctggcgg	caagaagcc	3300
atccagttt	cttgcaggg	cttcccaacc	ttaccagagg	gcgccttc	tggcaattcc	3360
ggttcgcttgc	ctgtccatcaa	aaccgcggcag	tctactatc	gcccattgttgc	cccactgca	3420
gctacccgt	ttctctttgc	gcttgcgtt	tccctgttcc	agatagccca	gtagctgaca	3480
ttcatccggg	gtcagcaccg	tttctgcggg	ctggcttct	acgtgttccg	tttccttttag	3540
cagcccttgc	gccctgatgt	cttgcggcag	cgtgaagcta	attcatgtt	aaattttgt	3600
taaatcagct	catttttaa	ccaataggcc	gaaatcgca	aaatccctta	taaatcaaaa	3660
gaatagcccg	agataggggtt	gagtgttgc	ccagtttgc	acaagagtcc	actattaaag	3720
aacgtggact	ccaacgtcaa	agggcgaaaa	accgtctatc	agggcgatgg	ccggatcagc	3780
ttatgcggtg	tgaataccg	cacagatgcg	taaggagaaa	ataccgcattc	aggcgctctt	3840
ccgcttcctc	gctcaactgac	tcgctgcgt	cggtcgatcg	gctgcggcga	gcggtatcag	3900
ctcactcaaa	ggcgttaata	cggttatcca	cagaatcagg	ggataacgc	ggaaagaaca	3960
tgtgagcaaa	aggccagcaa	aaggccaggg	accgtaaaaa	ggccgcgttgc	ctggcgtttt	4020
tccataggct	ccgcggccct	gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	4080
gaaacccgac	aggactataa	agataccagg	cgtttcccc	tggaaatcttcc	ctcgatcgct	4140
ctcctgttcc	gaccctgccc	cttaccggat	acctgtccgc	cttctccct	tcgggaagcg	4200
tggcgcttgc	tcatagctca	cgctgttaggt	atctcagttc	ggtgttaggtc	gttcgccttca	4260
agctgggctg	tgtgcacgaa	ccccccgttc	agcccgaccg	ctgcgcctt	tccggtaact	4320
atcgcttgc	gtcccaaccg	gtaagacacg	acttatcgcc	actggcagca	gccactggta	4380
acaggattag	cagagcgagg	tatgttggcg	gtgctacaga	gttcttgcag	ttgtggcccta	4440
actacggcta	cactagaagg	acagtatttg	gtatctcgcc	tctgttgcag	ccagttacct	4500
tcggaaaaag	agtttgttgc	tcttgcgttgc	gcaaacaac	caccgcgttgc	agcgggtgggt	4560
ttttgttttgc	caagcagcag	attacgcgc	gaaaaaaaaa	atctcaagaa	gatccctttga	4620
tctttcttac	tgaacggtga	ccccacccgg	aatttgcg			4657

<210> 52
<211> 4774
<212> DNA
<213> Artificial Sequence

<220>
<223> pCMVKm2.pCAP5opt HML-2 vector

<400> 52						
gcccgaaat	ttcgactcta	ggccatttgc	tacgttgc	ctatatcata	atatgtacat	60
ttatatttgc	tcatgtccaa	tatgaccg	atgttgc	atgttattga	ctagttatta	120
atagtaatca	attacgggg	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggccgc	ctggctgc	gcccaacgc	ccccgcctt	tgacgtcaat	240
aatgacgtat	gttccatag	taacgcaat	agggactt	cattgacgtc	aatgggtgg	300
gtatattacgg	taaaactgccc	acttggc	acatcaagtg	tatcatatgc	caagtcgc	360
ccctatttgc	gtcaatgac	gtaaatggcc	cgccctggc	tatgccc	acatgac	420
acgggacttt	cctacttgc	agtacatcta	cgtttagtgc	atcgctt	ccatgggtat	480
gccccgggg	cgtacacca	atgggcgttgc	atagcggtt	gactca	gatttcaag	540
tctccaccc	attacgc	atggagttt	gttttggc	caaaatcaac	gggacttcc	600
aaaatgtcg	aataaccccg	ccccgttgc	gcaaata	ggtaggcgt	tacgggtgg	660
gtctatata	agcagagctc	gttttagtgc	ccgtcagatc	gcctggagac	ccatccac	720

Substitute Sequence Listing_USSN 10587032_PP019482.007

ctgtttgac	ctccatagaa	gacaccggga	ccgatccagc	ctccgcggcc	gggaacggtg	780
cattgaaacg	cggattcccc	gtgccaagag	tgacgtaagt	accgcctata	gactctatag	840
gcacacccct	ttggctctta	tgcatgctat	actgttttgc	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataagg	gatggtata	cttagccat	aggtgtgggt	tattgaccat	960
tattgaccac	cccccttattt	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctcttattt	gctatatgcc	aataactctgt	ccttcagaga	ctgacacgga	1080
ctctgtatTTT	ttacaggat	gggtcccatt	tattatttac	aaattcacat	atacaacaac	1140
gccgtcccc	gtgcccgcag	ttttattaa	acatagcgtg	ggatctccac	gcaaatctcg	1200
ggtagtgtt	ccggacatgg	gctctctcc	ggtagcggcg	gagttccac	atccgagccc	1260
tggtcccatt	cctccagcg	ctcatggtc	ctcggcagct	ccttgtctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccatgt	tgccgcacaa	ggccgtggcg	1380
gttaggtatg	tgtctaaaa	tgagctcgga	gattgggctc	gcaccgcgt	cgcagatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgtatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgcgttta	acggtgagg	gcagtgtagt	ctgagcagta	1560
ctcggtctg	ccgcgcgcgc	caccagacat	aatacgatgc	agactaacag	actgttccctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgcacc	atgaacccca	gcgagatgca	1680
gwgcaaggcc	cccccccgcc	gccggcgc	ccgcaaccgc	gccccctga	cccacaagat	1740
gaacaagatg	gtgacagcg	aggagcagat	gaagtcggcc	agcacaaga	aggccgagcc	1800
ccccacattgg	gcccagctga	agaagctgac	ccagctgccc	accaagtacc	tggagaacac	1860
caagggtgacc	cagacccccc	agagcatgt	gctggccgccc	ctgatgatcg	tgagcatggt	1920
ggtgttacccc	accgcggccca	agcggcagcg	ccccagccgc	acggccacg	acgacgacgg	1980
cggcttcgtg	gagaagaagc	ggggcaagtg	ccggcgtc	ttctgtctgt	gcgactgcta	2040
ctgcgtgtgc	gtggagcga	ggccgcacc	ccgcctgcac	ttctgtctgt	acgcttaaag	2100
aattcagact	cgagcaagtc	tagaaagcc	tggatatcgg	atccactacg	cgttagagct	2160
cgctgtatcg	cctcgactgt	gccttctgt	tgccagccat	ctgttgttt	cccctcccc	2220
gtgccttcct	tgacccttgg	aggtgcact	cccaactgtcc	tttcttaata	aaatgaggaa	2280
attgcatcg	attgtctgag	taggtgtcat	tctattctgg	gggggtgggt	ggggcaggac	2340
agcaaggggg	aggattggga	agacaatagc	aggggggtgg	gcaagaact	ccagcatgag	2400
atccccgcgc	tggaggatca	tccagccggc	gtcccgaaaa	acgattccga	agcccaacct	2460
ttcatagaa	gcggcgggtgg	aatcgaaatc	tcgtgtatgc	aggttggcg	tcgcttggtc	2520
ggtcatttcg	aaccccgag	tcccgctcag	aagaactcgt	caagaaggcg	atagaaggcg	2580
atgcgtcg	aatcgggagc	ggcgatacc	taaagcacga	ggaagcggc	agcccattcg	2640
ccgccaagct	cttcagcaat	atcacgggt	gccaacgct	tgtcctgata	gcggtccgccc	2700
acaccagcc	ggccacagtc	gatgaatcca	gaaaagccgc	cattttccac	catgatattc	2760
ggcaagcagg	catcgccatg	ggtcacgc	agatctcgc	cgtcgggc	gcgcgccttg	2820
agcctggcga	acagttcggc	ttgcgcgagc	ccctgtatgc	cttcgtccag	atcatctga	2880
tcgacaagac	cggcttcatt	ccgagatcgt	gctcgctcg	tgcgatgtt	cgcttggtgg	2940
tcgaatggc	aggtagccgg	atcaagcgta	tgccagccgc	gcattgcate	agccatgatg	3000
gatactttct	ccgcaggagc	aaaggtagat	gacaggagat	cctgccccgg	cacttcgccc	3060
aatagcagcc	agtcccttc	cgcttcgt	acaacgtcga	gcacagctgc	gcaaggaaacg	3120
cccgctgtgg	ccagccacga	tagccgcgt	gcctcgct	gcagttcatt	cagggcaccg	3180
gacagggtcg	tcttgacaaa	agaacccggg	cgccctgtcg	ctgacagccg	gaacacggcg	3240
gcatcagagc	agccgattgt	ctgttgtg	cagtcatgc	cgaatagcc	ctccacccaa	3300
gcggccggag	aacctcggt	caatccatct	tgttcaatca	tgccaaacga	tcctcatcct	3360
gtctcttgc	cagatcttgc	tccctcg	catcagatcc	ttggcggcaa	gaaagccatc	3420
cagtttactt	tgccagggtt	cccaac	ccagaggcg	ccccagctgg	caattccgt	3480
tcgcttgcgt	tccataaaaac	cgcccagtct	agctatcgcc	atgtaaagccc	actgcaagct	3540
acctgtttc	tctttcg	tgcg	tttgtccaga	tagcccagta	gctgacattc	3600
atccgggtc	agcaccgtt	ctgcggact	gctttctac	tgttccgc	cctttagcag	3660
cccttgcgc	ctgagtgc	gcggcagcgt	gaagctaatt	catggtaaa	tttttggtaa	3720
atcagctcat	tttttaacca	ataggccaa	atccggaaaa	tcccttataa	atcaaagaa	3780
tagccgaga	tagggttgg	tgttgc	gttggaaaca	agagtccact	attaaagaac	3840
gtggactcca	acgtcaaaagg	gcgaaaaacc	gtctatcagg	gcgatggccg	gatcagctta	3900
tgcgggtgt	aataccgcac	agatgcgtaa	ggagaaaata	ccgcatcagg	cgctctccg	3960
cttcctcgct	cactgactcg	ctgcgtcg	tcgttgcgt	gcggcgagcg	gtatcagctc	4020
actcaaaggc	ggtataacgg	ttatccacag	aatcagggg	taacgcagga	aagaacatgt	4080
gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	cgcgttgc	gcgtttttcc	4140
ataggctccg	ccccccctgc	gagcatcaca	aaaatcgac	ctcaagtcgt	aggtggcgaa	4200
acccgacagg	actataaaga	taccaggcg	ttccccctgg	aagctccctc	gtgcgtctc	4260
ctgttccgac	cctggcgctt	accggatacc	tgtccgc	tctccctcg	ggaagcgtgg	4320
cgcttctca	tagtcacgc	tgttaggtt	tcagttcggt	gtaggtcg	cgctccaagc	4380
tgggctgtt	gcacgaaccc	cccggttc	ccgaccgcgt	cgccttatcc	ggtactatc	4440
gtcttgatc	caacccggta	agacacgact	tatcgccact	ggcagcagcc	actggtaaca	4500

Substitute Sequence Listing_USSN 10587032_PP019482.007

ggatttagcag	agcgaggat	gtaggcggtg	ctacagagtt	cttgaagtgg	tggcctaact	4560
acggctcac	tagaaggaca	gtatTTggta	tctgcgc	gctgaagcca	gttaccc	4620
aaaaaagagt	ttgttagctct	tgatCCggca	aacaaccac	cgctggtagc	ggtgttttt	4680
ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaggatc	tcaagaagat	ccttgatct	4740
tttctactga	acggtgatcc	ccacccgaaat	tgcg			4774

<210> 53
<211> 6483
<212> DNA
<213> Artificial Sequence

<220>
<223> pCMVKm2.gag wt PCAV vector

<400> 53	ttcgactcta	ggccattgca	tacgttgat	ctatata	atatgtacat	60
ttatatttgc	tcatgtccaa	tatgaccgccc	atgttgacat	tgattattga	ctagttatta	120
atagtaatca	attacgggg	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggcccgc	ctggctgacc	gcccaacgac	ccccggccat	tgacgtcaat	240
aatgacgtat	gttcccata	taacgccta	aggggacttc	cattgacgtc	aatgggttgg	300
gtatTTacgg	taaactgccc	acttggcagt	acatcaagt	tatcatatgc	caagtccgccc	360
ccctatttgc	gtcaatgacg	gttaatggcc	cgcctggcc	tatgccc	acatgacc	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgcttata	ccatgggtat	480
gcggttttgg	cagtacacca	atggggcgtgg	atagcggtt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atggggagttt	gtttggcac	caaaatcaac	gggacttcc	600
aaaatgtcgt	aataaccccg	ccccgttgc	gcaaatgggc	ggtagggcgtg	tacgggtgg	660
ggtctatata	agcagagctc	gttttagtga	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgc	ctccatagaa	gacaccggga	ccgatcc	ctccgcggcc	ggaaacgg	780
cattggAACG	cggattcccc	gtgccaagag	tgacgtaa	accgcctata	gactctatag	840
gcacacccct	ttggctctta	tgcatgttat	actgttttgc	gcttggggcc	tatacacccc	900
cgcttcc	tgctataggt	gatggtata	cttagccat	aggtgtgggt	tattgaccat	960
tattgaccac	tcccctattt	gtgacgatac	tttccattac	taatccat	catggctt	1020
tgccacaact	atctctattt	gctatatgc	aataactctgt	ccttcagaga	ctgacacgg	1080
ctctgtat	ttacaggat	gggtcccatt	tattatttac	aaatttacat	ataacaaca	1140
gccgtcccc	gtgcccgc	tttttattaa	acatagcgt	ggatctccac	gcgaatctcg	1200
ggtacgtt	ccggacatgg	gctttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tggtccc	cccccagcg	ctcatggt	ctcggcagct	ccttgctct	aacagtgg	1320
gccagactt	ggcacagcac	aatgcccacc	accaccagtg	tgccgcacaa	ggccgtggcg	1380
gtagggat	tgtctaa	tgagctcg	gattgggctc	gcaccgctg	cgcagatgg	1440
agacttaagg	cagcggcaga	agaagatgc	ggcagctgag	ttgtgtt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgtgtt	acgggtgggg	gcagtgtat	ctgagcgt	1560
ctcggtct	ccgcgcgcgc	caccagat	aatagctgac	agactaacag	actgttcc	1620
tccatgggtc	tttctgc	tcaccgtc	cgacgc	atggggcaaa	ctgaaagtaa	1680
atatgcct	tatctcagct	ttttaaaat	tcttttaa	agaggggggg	ttagagctt	1740
tacagaaaat	ctaattacgc	tatTTcaaa	aatagaacaa	ttctgccc	gtttccaga	1800
acagggaaact	tttagatctaa	aagattggg	aaaaatttgc	aaagaattaa	aacaagca	1860
tagggaaaggt	aaaatcatcc	cacttac	atggaat	tggccat	ttaaagca	1920
tttagaaacca	tttcaa	gagaagat	ttttcagtt	tctgtgccc	ctaaaagct	1980
tgtacagat	tgtgaaga	aggcagg	agaatccc	caaggaacgg	aaagttcaca	2040
ttgtaaat	gtacgtat	ctgtat	tcagtca	aaaaatgtt	actacagt	2100
attacaggag	ataatatacc	ctgaat	aaaattggg	gaaggagg	cagaat	2160
ggggccatca	gaggctaa	cacgat	atcaact	cctccgt	ttcagat	2220
tgtacatt	caacct	cgcagg	acaagc	accca	aaaatca	2280
agaaagg	agagtct	tcccgg	gccaact	atacagt	cacaat	2340
gccggtagaa	aataagaccc	aaccgt	agtttata	tacggc	caaccg	2400
tcagtatcg	ccctcc	aggttca	cagac	cgtgt	ctgtcc	2460
tagcacggc	ccatacc	aacc	gatggc	aattc	caacac	2520
cgcggcgt	tatcc	cggcc	gagactt	cctac	cacgt	2580
acagggtt	gca	cgtt	tgaagcc	aaacagg	atctt	2640
atggcggtt	ctggta	tacaact	acaggccc	ggagag	aaggtag	2700
gcctgccc	gctg	gatgt	tttccat	aaaatgtt	aaaatataa	2760
ggaaggagtt	aaacaat	actc	ccttata	agaacat	tagattt	2820
tgctcatg	aatagactt	ctcc	cttggaaatt	ttggccaaat	ttccctt	2880

Substitute Sequence Listing_USSN 10587032_PP019482.007

atccctctcg	tatctacagt	ttaaaacctg	gtggatttat	ggagtacaag	aacaggtagc	2940
aaaaaatcg	gctactaagc	ccactgttaa	tatagacgc	gaccaattgt	taggaacagg	3000
tccaaattgg	agcaccatta	accaacaatc	agtatgcag	aataggctt	ttgaacaagt	3060
aaggctatt	tgcctcagg	cctggggaaa	aattcagagc	ccaggaacag	cttccctat	3120
taattcaatt	agacaaggct	ctaaagagcc	atatcctgac	tttgtggcaa	gattacaaga	3180
tgctgtcaa	aagtcttata	cagatgacaa	tgcccgaaaa	gttattgtat	aattaatggc	3240
ctatgaaaat	gcaaatccag	aatgtcagtc	ggccataaaag	ccattaaaag	gaaaagttcc	3300
agcaggagtt	gatgtatata	cagaatatgt	gaaggcttgt	gatgggattt	gaggagctat	3360
gcataaggca	atgctaattgg	ctcaagcaat	gagggggctc	actctaggag	gacaaggtag	3420
aacatttggg	aaaaaatgtt	ataattgtgg	tcaaattcggt	catctgaaaaa	ggagttgcc	3480
agtcttaat	aaacagaata	taataaaatca	agctattaca	gcaaaaaata	aaaagccatc	3540
tggcctgtgt	ccaaaatgtg	gaaaaggaaa	acattggcc	aatcaatgtc	attctaaatt	3600
tgataaggat	gggcaaccat	tgtcgggaaa	caggaagagg	ggccagccctc	aggccccca	3660
acaaactggg	gcattcccg	ttcaactgtt	tgccctctcg	ggtttcaag	gacaacaacc	3720
cctacagaaa	ataccaccac	ttcagggagt	cagccaatta	caacaatcca	acagctgtcc	3780
cgcgcacag	caggcagcac	cgcagtaaga	attcagactc	gagcaagtct	agaaagccat	3840
ggatatcgg	tccactacgc	gttagagctc	gctgatcagc	ctcgactgtg	ccttctagtt	3900
gccaggccatc	tgtgttttc	ccctccccc	tgcccttcctt	gaccctggaa	gtgtccactc	3960
ccactgtcct	ttcctaataa	aatgaggaaa	ttgcatcgca	ttgtctagt	aggtgtcatt	4020
ctattctggg	gggtgggggt	gggcaggaca	gcaaggggga	ggattgggaa	gacaatagca	4080
gggggggtggg	cgaagaactc	cagcatgaga	tcccccgcct	ggaggatcat	ccagccggcg	4140
tcccggaaaaa	cgattccgaa	gccccacctt	tcatagaagg	cggcgggtgg	atcgaatct	4200
cgtgatggca	ggttgggcgt	cgcttggtc	gtcatttcga	accccagagt	cccgctcaga	4260
agaactcgtc	aagaaggcga	tagaaggcga	tgcgctgcga	atcgggagcg	gcgataccgt	4320
aaagcacgag	gaagcggcga	gcccattcgc	cgccaagctc	ttcagcaata	tcacggtag	4380
ccaacgctat	gtcctgatag	cggtccgca	cacccagccg	gccacagtct	atgaattccag	4440
aaaagcggcc	attttccacc	atgatattcg	gcaagcaggc	atcgccatgg	gtcacgacga	4500
gatcctcgcc	gtcgggcatt	cgcccttga	gcctggcgaa	cagttcggt	ggcgcgagcc	4560
cctgatgctc	ttcgtccaga	tcatccgtat	cgacaagacc	ggcttccatc	cgagtacgt	4620
ctcgctcgat	gcgatgtttc	gcttgggtgt	cgaatggca	gttagccgga	tcaagcgtat	4680
gcagccgccc	cattgcatca	gccatgatgg	atacttttc	ggcaggagca	aggtgagatg	4740
acaggagatc	ctgccccggc	acttcgccc	atagcagcca	gtcccttccc	gtttcagtga	4800
caacgtcgag	cacagctcg	caaggaacgc	ccgtcgtggc	cagccacgt	agccgcgt	4860
cctcgctctg	cagttcattt	agggcaccgg	acaggtcggt	tttgacaaaaa	agaaccgggc	4920
gccctcgcc	tgacagccgg	aacacggcg	catcagagca	gccgattgtc	tggtgtccc	4980
agtcatagcc	gaatagcctc	tccacccaag	cgcccgagaa	acctgcgtgc	aatccatctt	5040
gttcaatcat	gcgaaaacgt	cctcatctg	tctcttgcatt	agatcttgcatt	ccccctgcgc	5100
atcagatcct	tggcggcaag	aaagccatcc	agtttacttt	gcagggcttc	ccaaacctac	5160
cagaggcgc	cccagctggc	aattccgggtt	cgctgtgt	ccataaaaaacc	gcccagtcata	5220
gctatcgcca	tgtaaagccca	ctgcaagct	cctgttttc	cttgcgtt	gcgtttccc	5280
ttgtccagat	agcccagtat	ctgacattca	tccgggggtca	gcaccgtttc	tgcggactgg	5340
ctttctacgt	gttccgcttc	ctttagcagc	ccttgcgtcc	tgagtgttt	cggcagcgt	5400
aagctaattc	atggttaaat	tttttttttt	tcagtcatt	tttttaccaa	tagggccaaa	5460
tcggcaaaat	cccttataaa	tcaaaaagaat	agcccagat	agggttgagt	gttgtttccag	5520
tttggaaaca	gagtccacta	ttaaaaagacg	tggactccaa	cgtcaaaagg	cgaaaaaccg	5580
tctatcaggg	cgatggccgg	atcagtttat	gcgggtgtaa	ataccgcaca	gatgcgtaa	5640
gagaaaatac	cgcatcaggc	gctctccgc	ttcctcgtct	actgactcgc	tcgcgtcggt	5700
cgttcggtcg	cggcgagcgg	tatcagctca	ctcaaaaggcg	gtaatacgt	tatccacaga	5760
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaaggc	cagaaaaagg	ccaggaaccg	5820
taaaaaggcc	gcgttgcgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	5880
aaatcgacgc	tcaagtcaaa	ggtggcggaa	cccgcacagga	ctataaaagat	accaggcggt	5940
tccccctgg	agctccctcg	tgcgtctcc	tgttccgacc	ctggcgctt	ccggatacc	6000
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctat	agctcagct	gttagtatct	6060
cagttcggtg	taggtcgatc	gctccaaagct	gggctgtgt	cacgaacccc	cggttcagcc	6120
cgaccgctgc	gccttatcc	gtaactatcg	tcttgcgtcc	aacccggtaa	gacacgactt	6180
atcgccactg	gcagcagcc	ctggtaacag	gattagcaga	gcgaggatgt	taggcgtgc	6240
tacagagttc	ttgaagtgtt	ggcctaacta	cggttacact	agaaggacag	tatttggtat	6300
ctgcgtctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagcttt	gatccggcaa	6360
acaaaccacc	gctggtagcg	gtgggttttt	tttttgcaag	cagcagatta	cgccgagaaa	6420
aaaaggatct	caagaagatc	ctttgatctt	ttctactgaa	cggtgatccc	caccggaaatt	6480
	gcg.					6483

Substitute Sequence Listing_USSN 10587032_PP019482.007

<211> 6340

<212> DNA

<213> Artificial Sequence

<220>

<223> pCMVKm2.gagopt HML-2 vector

<400> 54

gccgcggaat	ttcgactcta	ggcattgca	tacgttgtat	ctatatacata	atatgtacat	60
ttatattggc	tcatgtccaa	tatgaccgcc	atgttgacat	tgattattga	ctagttatta	120
atagaatca	attacggggt	cattagttca	tagccccat	atggagttcc	gcgttacata	180
acttacggta	aatggcccgc	ctggctgacc	gccccaaacgac	ccccggccat	tgacgtcaat	240
aatgacgtat	gttcccatag	taacgccaat	agggactttc	cattgacgtc	aatgggtgga	300
gtatttacgg	taaactgccc	acttggcagt	acatcaagt	tatcatatgc	caagtccgccc	360
ccctatttgc	gtcaatgacg	gtaaaatggcc	cgcctggcat	tatgcccagt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtatttagtc	atcgcttatta	ccatgggtat	480
gcccgtttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atgggaggtt	gttttggcac	caaaatcaac	gggactttcc	600
aaaatgtcg	aataaccccg	ccccgttgac	gcaaaatgggc	ggtaggcgtg	tacggggga	660
ggtctatata	agcagagctc	gttttagtga	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgc	ctccatagaa	gacaccggga	ccgatccacg	ctccgcggcc	gggaacgggt	780
cattgaaacg	cggattccccc	gtgccaagag	tgacgttaat	accgcctata	gactctatag	840
gcacacccct	ttggctctta	tgcatgctat	actgttttgc	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataggt	gatggtata	cttagccat	aggtgtgggt	tattgaccat	960
tattgaccac	tccccttattt	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctcttattt	gctatatgcc	aataactctgt	ccttcagaga	ctgacacgg	1080
ctctgtat	ttacaggat	gggtcccatt	tattatttac	aaatttacat	ataacaacaac	1140
gccgtcccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gcaaatctcg	1200
ggtacgttt	ccggacatgg	gctttctcc	ggtagcggcg	gagttccac	atccgagccc	1260
tggtcccatt	cctccagcg	ctcatggtcg	ctcggcagct	ccttgcctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccatg	tgccgcacaa	ggccgtggcg	1380
gtagggat	tgtctaaaaa	tgagctcgga	gattgggctc	gcaccgctg	cgcagatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgtatt	ctgataaagag	1500
tcagaggtaa	ctcccgttgc	ggtgcgttta	acgggtggagg	gcagtgtatg	ctgagcagta	1560
ctcggtgct	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttccctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgcacc	atgggccaga	ccaagagcaa	1680
gatcaagac	aagtacgcca	gctacctgag	cttcatcaag	atctctgtg	agcgcggcgg	1740
cgtgaagg	agcaccaaaga	acctgatcaa	gctgttccag	atcatcgagc	agttctgccc	1800
ctggttcccc	gaggcaggca	ccctggacat	gaaggactgg	aagcgcatcg	gcaaggagct	1860
gaagcaggcc	ggccgcaga	gcaacatcat	ccccctgacc	gtgtggaaac	actgggcat	1920
catcaaggcc	gccctggagc	ccttccagac	cgaggagac	agcgtgagcg	tgagcagcgc	1980
ccccggcagc	tgcatcatcg	actgcaacga	gaacaccgc	aagaagagcc	agaaggagac	2040
cgagggcctg	cactgcgagt	acgtggccga	gcccgtatg	gcccagagac	cccagaacgt	2100
ggactacaac	cagctgcagg	aggtgatcta	ccccgagacc	ctgaagctgg	agggcaaggg	2160
ccccgagctg	gtggggcccca	gcgagagcaa	gccccgcggc	accagcccc	tgcccggcgg	2220
ccaggtgccc	gtgaccctgc	agccccagaa	gcaggtgaag	gagaacaaga	cccagcccc	2280
cgtggctac	cagtacttggc	ccccccgcga	gctgcagatc	cgcccccccc	ccgagagcc	2340
gtacggctac	cccgccatgc	ccccccccc	ccagggccgc	gccccctacc	cccagcccc	2400
cacccgcccgc	ctgaacccca	ccgccccccc	cagccgcag	ggcagcaagc	tgcacgagat	2460
catcgacaag	agccgcaagg	agggcgacac	cgaggcttgg	catttcccc	tgacccttgg	2520
gcccatgccc	cccgccgagg	gcccggcagg	gggcgagccc	cccaccgtgg	aggccccta	2580
caagagctt	agcatcaaga	agctgaagga	catgaaggag	ggcgtgaagc	agtacggccc	2640
caacagcccc	tacatgcgca	ccctgtctgga	cagcatgc	cacggccacc	gcctgtatccc	2700
ctacgactgg	gagatcctgg	ccaagagcag	cctgagcccc	agccagttcc	tgcagttcaa	2760
gacctgttgg	atcgacggcg	tgcaggagca	ggtgcgcgc	aaccgcgc	ccaaccccc	2820
cgtgaacatc	gacggccgacc	agctgttggg	catggccag	aactggagca	ccatcagcc	2880
gcaggccctg	atgcagaacg	aggccatcg	gcagggtgcgc	gccatctg	tgcgcgcctg	2940
ggagaagatc	caggaccccc	gcagcacctg	ccccagctt	aacaccgtgc	gccaggggcag	3000
caaggagccc	taccccgact	tcgtggcccg	cctgcaggac	gtggcccaga	agagcatcg	3060
cgacgagaag	gcccgcagg	tgatgttgg	gctgtatggcc	tacgagaacg	ccaaccccc	3120
gtcccaagac	gcccgcagg	caaggttgg	caaggttggcc	gcccgcagg	acgtgatcag	3180
cgagtagctg	aaggcctgcg	acggcatacg	cgccgcattat	cacaaggcca	tgctgtatggc	3240
ccaggccatc	accggcgtgg	tgctggggcg	ccaggtgcgc	accttgcggcc	gcaagtgcata	3300

Substitute Sequence Listing_USSN 10587032_PP019482.007

caactgcggc	cagatcgccc	acctgaagaa	gaactgcccc	gtgctgaaca	agcagaacat	3360
caccatccag	gccaccacca	ccggccgcga	gccccccgac	ctgtgcccccc	gctgcaagaa	3420
gggcaagcac	tgggcccagcc	agtggccgcag	caagttcgac	aagaacggcc	agcccccgtag	3480
cggcaacgag	cagcgcggcc	agccccaggc	ccccccagcg	accggcgccct	tccccatcca	3540
gcccttcgtg	ccccaggggct	tccagggcca	gcagcccccc	ctgagccagg	tgttccaggg	3600
catcagccag	ctgccccagt	acaacaactg	ccccccccc	caggccgcgg	tgcagcaggc	3660
ttaaagaatt	cagactcgag	caagtctaga	aagccatgga	tatcgatcc	actacgcgtt	3720
agagctcgct	gatcagcctc	gactgtgcct	tctagttgc	agccatctgt	tgtttgc(cc)cc	3780
tccccgtgc	cttccttgcac	cctgaaagggt	gccactccca	ctgtcctttc	ctaataaaat	3840
gaggaaattg	catcgatttgc	tctgagtagg	tgtcatttta	ttctgggggg	tgggggtgggg	3900
caggacagca	aggggggagga	ttgggaagac	aatagcaggg	gggtggggcga	agaactccag	3960
catgagatcc	ccgcgcgtgg	ggatcatcca	gccggcgtcc	cggaaaacga	ttccgaagcc	4020
caaccttca	tagaaggcgg	cggtggaaatc	gaaatctcg	gatggcaggt	tgggcgtcgc	4080
ttggtcggtc	atttcgaacc	ccagactccc	gctcagaaga	actcgtcaag	aaggcgatag	4140
aaggcgatgc	gctgcgaatc	gggagcggcg	ataccgtaaa	gcacgaggaa	gcgggtcagcc	4200
cattcgcgc	caagctcttc	agcaaatatca	cgggttagcca	acgctatgtc	ctgatagcgg	4260
tccgcacac	ccagccggcc	acagtcgtg	aatccagaaa	agccgcatt	ttccaccatg	4320
atattcggca	agcaggcattc	gccatgggtc	acgacgagat	cctgcgcgtc	gggcgcgc	4380
gccttgcggcc	ttggcgaacag	ttcggctggc	gcgagccct	gatgccttc	gtccagatca	4440
tcctgtatcg	caagaccggc	ttccatccga	gtacgtgtc	gctcgatcgc	atgtttcgct	4500
tggtggtcg	atgggcagggt	agccggatc	acgtatgc	gccgcgcgt	tgcatacgcc	4560
atgatggata	ctttctcgcc	aggagcaagg	tgagatgaca	ggagatctg	ccccggcact	4620
tcgccaata	gcagccagtc	ccttcccgct	tcagtgacaa	cgtcgagcac	agtcgcgcaa	4680
ggaacgcgg	tcgtggccag	ccacgatagc	cgcgctgcct	cgtcctgcag	ttcattcagg	4740
gcacccggaca	ggtcggcttt	gacaaaaaga	accgggcgcc	cctgcgtga	cagccgaaac	4800
acggccgcatt	cagagcagcc	gattgtctgt	tgtgcccagt	catagccgaa	tagcctctcc	4860
acccaagcgg	ccggagaacc	tgcgtgcaat	ccatcttgc	caatcatgc	aaacgatcct	4920
catcctgtct	cttgcatacg	tcttgcattcc	ctgcgcattc	agatcctgg	cggcaagaaa	4980
gccatccagt	ttactttgca	gggcttccca	actttaccag	aggcgccccc	agctggcaat	5040
tccgggttgc	ttgctgtcca	taaaaccgcc	cagtctagct	atgcctatgt	aagcccactg	5100
caagctacct	gctttcttctt	tgcgttgcg	ttttcccttg	tccagatagc	ccagtagctg	5160
acattcatcc	ggggtcagca	ccgtttctgc	ggactggctt	tctacgtgtt	ccgcttccctt	5220
tagcagccct	tgcgccttgc	gtgcttgcgg	cagcgttaag	ctaattcatg	gttaaatttt	5280
tgttaaatca	gctcattttt	taaccaatag	gccgaaatcg	gaaaaatccc	ttataaatca	5340
aaagaatagc	ccgagatagg	gttgagtgtt	gttccagttt	ggaacaagag	tccactattt	5400
aagaacgtgg	actccaacgt	caaagggcga	aaaaccgtct	atcagggcga	tggccgatc	5460
agcttatgc	gtgtgaaata	ccgcacagat	gcgttaaggag	aaaataccgc	atcaggcgct	5520
cttccgcctt	ctcgctact	gactcgtgc	gctcggtcg	tcgctgcgg	cgagcgttat	5580
cagctcactt	aaaggcggta	atacggttat	ccacagaatac	aggggataac	gcagggaaaga	5640
acatgtgagc	aaaaggccag	caaaggcca	ggAACCGTAA	aaaggccgcg	ttgctggcgt	5700
ttttccatag	gttccgcggcc	cctgacgagc	atcacaaaaaa	tgcacgctca	agttaggt	5760
ggcgaaaccc	gacaggacta	taaagatacc	aggcgtttcc	cccttggaaagc	tccctcgatc	5820
gctctccgt	tccgaccctg	ccgcttaccg	gataccgtc	cgcccttctc	ccttcggaa	5880
gcgtggcgt	ttctcatagc	tcacgctgt	ggtatctcg	ttcggtgtag	gtcgttcgt	5940
ccaaagctgg	ctgtgtgcac	gaacccccc	ttcagcccg	ccgtcgcc	ttatccggta	6000
actatcgct	ttagtccaa	ccggtaagac	acgacttatac	gccactggca	gcagccactg	6060
gtaacaggat	tagcagagcg	aggtatgtag	gcggtgctac	agagttcttg	aagtgggtgc	6120
ctaactacgg	ctacactaga	aggacagtat	ttgttatctg	cgctctgct	aagccagtt	6180
ccttcggaaa	aagagtgtt	agctcttgc	ccggcaaaaca	aaccaccgt	ggtagcggtg	6240
gttttttgt	ttgcaaggcag	cagattacgc	gcagaaaaaa	aggatctcaa	gaagatccctt	6300
tgtatcttcc	tactgaacgg	tgtatccccac	cggaatttgc			6340

<210> 55
<211> 5344
<212> DNA
<213> Artificial Sequence

<220>
<223> pCMVKm2.Prototyp HML-2 vector

<400> 55
gccgcggaaat ttgcactcta ggccatttgc tacgttgtat ctatatcata atatgtacat 60
ttatattggc tcatgtccaa tatgaccgc atgttgacat tgattattga ctatgttatta 120

Substitute Sequence Listing_USSN 10587032_PP019482.007

atagtaatca	attacggggt	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggcccg	ctggctgacc	gccccaaacgac	ccccggccat	tgacgtcaat	240
aatgacgtat	gttcccata	taacgccaat	agggacttc	cattgacgtc	aatgggtgga	300
gtatttacgg	taaactgccc	acttggcagt	acatcaagt	tatccatatgc	caagtccgccc	360
ccctattgac	gtcaatgacg	gtaaatggcc	cgccctggcat	tatgcccagt	acatgaccctt	420
acgggacttt	cctacttggc	agtacatcta	cgtatttagtc	atcgcttatta	ccatgggtat	480
gcgggtttgg	cagtagcacca	atgggcgtgg	atagcggtt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atggaggtt	gtttggcac	caaaatcaac	gggactttcc	600
aaaatgtcgt	aataaccccg	ccccgttgac	gcaaataggc	ggtaggcgtg	tacgggtggaa	660
ggtctatata	agcagagctc	gtttagtgaa	cgctcagatc	gcctggagac	gccatccacg	720
ctgttttgac	ctccatagaa	gacaccggga	ccgatccagc	ctcccgccg	gggaacggtg	780
cattggAACG	cggattcccc	gtgccaagag	tgacgttaat	accgcctata	gactctatag	840
gcacacccct	ttggctctt	tgcatgctat	actgttttgc	gcttggggcc	tatacacccc	900
cgcttcctt	tgctataggt	gatggtata	cttagccat	agggtgggt	tattgaccat	960
tattgaccac	tcccctattt	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctctattt	gctatatgcc	aataactctgt	ccttcagaga	ctgacacgg	1080
ctctgtatTT	ttacaggat	gggtccattt	tattattatc	aaattcacat	atacaacaac	1140
ggcgtcccc	gtggcccg	tttttattaa	acatagcgt	ggatctccac	gcgaatctcg	1200
ggtacgttt	ccggacatgg	gctttctcc	ggtagcggcg	gagtttccac	atccgagccc	1260
tggtcccat	cttccagcg	ctcatggtc	ctcggcagct	ccttgcctt	aacagtggag	1320
gccagacttA	ggcacagcac	aatggccacc	accaccatgt	tgccgcacaa	ggccgtggcg	1380
gtagggatgt	tgtctaaaaa	tgagctcgga	gattgggttc	gcacccgtga	cgcagatgga	1440
agacttaagg	cagcgccaga	agaagatgca	ggcagctgag	ttgttgtatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgcgttta	acgggtggagg	gcagtgtat	ctgagcagta	1560
ctcggtctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttctt	1620
tccatgggtc	ttttctgcag	tcaccgtcg	cgacgcacc	atgtgggcca	ccatcggtgg	1680
caagcgcgc	aaggggcccg	ccagggccc	caccaccaac	tggggcatcc	ccaaacagcgc	1740
catctgcagc	agcggttca	gcggcaccac	caccccccac	gtgcccagcg	tgagcggcaa	1800
caagccgtg	accaccatcc	agcagctgag	ccccggccacc	agcggcagcg	ccgcccgtgga	1860
cctgtgcacc	atccaggccg	tgaggctgct	gcccggcgag	cccccccaaga	agaccccac	1920
cggcgtgtac	ggccccctgc	ccaaggggcac	cgtgggcctg	atcttgggca	gcagcagcct	1980
gaacctgaag	ggcgtgcaga	tccacaccag	cgtgggtggac	agcactaca	agggcgagat	2040
ccagctgggt	atcagcagca	gcatccccctg	gagcggccagc	ccccggcacc	gcatcgccca	2100
gctgctgtcg	ctgcccata	tcaagggcgg	caacagcgag	atcaagcgca	tcggcggcct	2160
gggcagcacc	gaccccccgg	gcaaggccgc	ctactgggccc	agccaggtga	gcgagaaccg	2220
ccccgtgtgc	aaggccatca	tccaggggcaa	gcagttcgag	ggccttggtg	acaccggcgc	2280
cgacgtgagc	atcatcgccc	tgaaccagt	gccccagaac	tggcccaagc	agaaggccgt	2340
gaccggcctg	gtgggcatcg	gcaccggccag	cgaggtgtac	cagacccacg	agatccgtca	2400
ctggcttggc	ccggacaaacc	aggagagcac	cgtgcagcccc	atgtacca	gcatccccct	2460
gaacctgtgg	ggccgcgacc	tgctcagca	gtggggccccc	gagatccacca	tgccgcggcc	2520
cagctacagc	cccaccagcc	agaagatcat	gaccaagat	ggctacatcc	ccggcaaggg	2580
cctgggcaag	aacgaggacg	gcatcaagat	ccccgtggag	gccaagatca	accaggagcg	2640
cgagggcattc	ggcaacccct	gchgctaaag	aattcagact	cgagcaagtc	tagaaagcca	2700
tggatatcg	atccactac	cgtttagagct	cgctgatcag	cctcgactgt	gccttcttagt	2760
tgccagccat	ctgttggttt	ccccctcccc	gtgccttcc	tgacccttgg	aggtgcact	2820
cccactgtcc	tttcctaata	aaatgaggaa	attgcattcg	attgtcttag	taggtgtcat	2880
tctattctgg	ggggtgggg	ggggcaggac	agcaagggggg	aggattggga	agacaatagc	2940
aggggggtgg	gcaagaact	ccagcatgag	atccccggcc	tggaggatca	tccagccggc	3000
gtccccggaa	acgattccga	agccccaaact	ttcatagaag	gcggcgggt	aatcgaaatc	3060
tcgtgtatgc	aggttgggc	tcgcttggtc	ggtcatttcg	aaccccaagag	tcccgtcag	3120
aagaactcgt	caagaaggcg	atagaaggcg	atgcgtcg	aatcgggagc	ggcgataccg	3180
taaagcacga	ggaagcggtc	agccccattcg	ccgccaagct	cttcagcaat	atcacgggt	3240
gccaacgcta	tgtcctgata	gcgggtccg	acaccacggc	ggccacagtc	gatgaatcca	3300
gaaaagccgc	cattttccac	catgatattc	ggcaagcagg	catcgccat	ggtcagcagc	3360
agatccgc	cgtcgggc	gcccgccttgc	agcctggcga	acagttcg	tggcgcgagc	3420
ccctgtatgc	cttcgtcc	atcatccgt	tcgacaagac	cggcttccat	ccgagttacgt	3480
gctcgtcg	tgcgatgttt	cgcttgggt	tcgaatggc	aggttagccgg	atcaagcgta	3540
tgcagccgc	gcatttgatc	agccatgtat	gatactttc	cggcaggagc	aaggtgagat	3600
gacaggagat	cctggccccc	cacttcgccc	aatagcagcc	agtcccttcc	cgcttca	3660
acaacgtcg	gcacagctgc	gcaaggaacg	cccgtcg	ccagccacga	tagccgcgt	3720
gcctcgatc	gcagttcatt	cagggcaccg	gacaggtcg	tcttgcacaa	aagaaccggg	3780
cgccctcg	ctgacagccg	gaacacggcg	gcatcagac	agccgattgt	ctgttgc	3840
cagtcatagc	cgaatagcct	ctccacccaa	gcggccggag	aacctgcgt	caatccatct	3900

Substitute Sequence Listing_USSN 10587032_PP019482.007

tgttcaatca	tgcgaaacga	tcctcatcct	gtcttttat	cagatcttgc	tccctgcgc	3960
catcagatcc	ttggcgccaa	gaaagccatc	cagtttactt	tgcaaggctt	cccaaccctt	4020
ccagagggcg	ccccagctgg	caattccgtt	tcgcttgcgt	tccataaaaac	cgccccagtct	4080
agctatcgcc	atgttaagccc	actgcaagct	acctgccttc	tctttgcgtt	tgcggtttcc	4140
cttgcctaga	tagccctagta	gctgacattc	atccggggtc	agcaccgttt	ctgcggactg	4200
gctttctacg	tgttcccgctt	ccttttagcag	cccttgcgcc	ctgagtgctt	gcggcagcgt	4260
gaagctaatt	catggtaaaa	tttttgttaa	atcagctcat	tttttaacca	ataggccgaa	4320
atcggcaaaa	tcccttataa	atcaaagaa	tagcccggaa	tagggtttag	tgttgttcca	4380
gtttgaaaca	agagtccact	attaaagaac	gtggacttca	acgtcaaagg	gcgaaaaacc	4440
gtctatcagg	gcgatggccg	gatcagctt	tgccgtgtga	aataccgcac	agatgcgtaa	4500
ggagaaaaata	ccgcatcagg	cgcttcccg	cttcctcgct	cactgactcg	ctgcgctcgg	4560
tcgttcggct	gccccggcg	gtatcagctc	actcaaaggc	ggttaatacgg	ttatccacag	4620
aatcagggga	taacgcagga	aagaacatgt	gagcaaaagg	ccagcaaaag	gccaggaacc	4680
gtaaaaaggc	cgcgttgctg	gcgttttcc	ataggctccg	ccccccgtac	gagcatcaca	4740
aaaatcgcac	ctcaagtca	aggtggcgaa	acccgcacagg	actataaaga	taccaggcgt	4800
ttccccctgg	aagctccctc	gtgcgtctc	ctgttccgac	cctgcccgtt	accggatacc	4860
tgtccgcctt	tctcccttcg	ggaagcgtgg	cgctttctca	tagctcacgc	tgttaggtatc	4920
tcagttcggt	gttaggtcggt	cgcttcaagg	tgggctgtgt	gcacgaaccc	cccggtttagc	4980
ccgaccgcgt	cgccttattcc	ggtaactatc	gtcttgagtc	caacccggta	agacacgact	5040
tatcgccact	ggcagcggcc	actggtaaca	ggatttagcag	agcggatgt	gtaggccgtg	5100
ctacagagtt	cttgaagttgg	ttggcttaact	acggctacac	tagaaggaca	gtatttggta	5160
tctgcgtct	gctgaagcca	gttaccttcg	gaaaaagagt	tggtagctt	tgatccggca	5220
aacaaaccac	cgctggtagc	ggtggttttt	ttgtttgc	gcagcagatt	acgcgcagaa	5280
aaaaaggatc	tcaagaagat	cctttgatct	tttctactga	acggtgatcc	ccaccggaaat	5340
						5344

<210> 56
<211> 7211

<212> DNA

<213> Artificial Sequence

<220>
<223> pCMVKm2.Polopt HML-2 vector

<400> 56						
gccgcggaat	ttcgactcta	ggccatttgc	tacgttgtat	ctatatcata	atatgtacat	60
ttatatttgc	tcatgtccaa	tatgaccgccc	atgttgcacat	tgattattgt	ctagttatta	120
atagtaatca	attacgggg	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggccgc	ctggctgacc	gccccaaacgac	ccccggccat	tgacgtcaat	240
aatgacgtat	gttcccatag	taacgcctaa	aggggacttc	cattgacgtc	aatgggttgg	300
gtatttacgg	taaactggcc	acttggcagt	acatcaagt	tatcatatgc	caagtccgccc	360
ccctatttgc	gtcaatgacg	gttaaatggcc	cgccctggcat	tatggccagt	acatgacatt	420
acgggacttt	cctacttggc	agtacatcta	cgtttagtgc	atcgcttata	ccatgggtat	480
gcgggtttgg	cagtacacca	atggggcgtgg	atagcggtt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atggggagttt	gttttggcac	caaaaatcaac	gggactttcc	600
aaaatgtcgt	aataaccccg	ccccgttgac	gcaaaatgggc	ggttaggcgtg	tacgggtgg	660
ggtctatata	agcagagctc	gttttagtga	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgc	ctccatagaa	gacaccggga	ccgatccagc	ctccggcgcc	ggaaacggtg	780
cattggaaacg	cggttccccc	gtgccaagag	tgacgttaat	accgcctata	gactctatag	840
gcacacccct	ttggctctt	tgcatgtat	actgttttg	gcttggggcc	tatacacccc	900
cgcttccat	tgctataggt	gatggtata	cttagccat	aggtgtgggt	tattgaccat	960
tattgaccac	tcccctattt	gtgacgatac	tttccattac	taatccat	catggctt	1020
tgccacaact	atctctattt	gctatatgc	aataactctgt	ccttcagaga	ctgacacgg	1080
ctctgtat	ttacaggat	gggtcccatt	tattatttac	aaatttacat	atacaacaac	1140
gccgtcccc	gtgccccgg	tttttattaa	acatagcgt	ggatctccac	gcgaatctcg	1200
ggtacgtgtt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tggtcccat	cctccagcg	ctcatggtgc	ctcggcagct	ccttgcctt	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagtg	tgccgcacaa	ggccgtggcg	1380
gtagggat	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcagatgg	1440
agacttaagg	cagcggcaga	agaagatgc	ggcagctgag	ttgttgtatt	ctgataagag	1500
tcagaggtaa	ctccccgttgc	ggtgtgtt	acgggtggagg	gcagtgtat	ctgagcagta	1560
ctcggtct	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgtttcctt	1620
tccatgggtc	ttttctgtcgt	tcaccgtcgt	cgacgcacc	atgaacaaga	gccgcagcgc	1680

Substitute Sequence Listing_USSN 10587032_PP019482.007

ccgcaaccgc	gagagcctgc	tgggcgccgc	caccgtggag	ccccccaagc	ccatccccct	1740
gacctggaag	accgagaagc	ccgtgtgggt	gaaccagtgg	cccctgccc	agcagaagct	1800
ggaggccctg	cacctgctgg	ccaacgagca	gctggagaag	ggccacatcg	agcccagctt	1860
cagccccctg	aacagccccg	tgttcgtat	ccagaagaag	agcggcaagt	ggcgcatgct	1920
gaccgacctg	cgcgcccgtga	acgcccgtat	ccagcccatg	ggccccctgc	agcccggcct	1980
gcccagcccc	gccatgatcc	ccaaggactg	gcccctgatc	atcatcgacc	tgaaggactg	2040
cttcttcacc	atccccctgg	ccgagcagga	ctgcgagaag	ttgccttca	ccatccccgc	2100
catcaacaac	aaggagccg	ccacccgctt	ccagtggaaag	gtgtgcccc	agggcatgct	2160
gaacagcccc	accatctgcc	agaccctcg	gggcccgc	ctgcagcccg	tgcgcgagaa	2220
gttcagcgcac	tgctacatca	tccactgcat	cgacgacatc	ctgtgcgccc	ccgagaccaa	2280
ggacaagctg	atcgactgct	acacccttc	gcaggccgag	gtggccaacg	ccggccctggc	2340
catcgccagc	gacaagatcc	agaccagcac	ccccttccac	tacctgggca	tgcagatcga	2400
gaaccgcaag	atcaagcccc	agaagatcga	gatccgcaag	gacaccctga	agaccctgaa	2460
cgacttccag	aagctgctgg	gcgacatcaa	ctggatccgc	cccaccctgg	gcatccccac	2520
ctacgcccatt	agcaacctgt	tcagcatcct	gcgcggcgcac	agcgcacctga	acagcaagcg	2580
catgctgacc	cccgaggcc	ccaaggagat	caagctggtg	gaggagaaga	tccagagcgc	2640
ccagatcaac	cgcacatgacc	ccctggcccc	cctgcagctg	ctgatcttc	ccaccgccc	2700
cagccccacc	ggcatcatca	tccagaacac	cgacctggtg	gatggagct	tcctgcccc	2760
cagcaccgtg	aagacccctca	ccctgtacct	ggaccagatc	gccaccctga	tcggccagac	2820
ccgcctgcgc	atcatcaagc	tgtgcggcaa	cgaccccgac	aagatctgg	tgccccctgac	2880
caaggagcag	gtgcgcgg	ccttcatcaa	cagcggcgc	tgaaagatcg	gcctggccaa	2940
cttcgtggcc	atcatcgaca	accatacc	caagaccaag	atcttccagt	tccngaagct	3000
gaccacctgg	atcctgccc	agatcacc	ccgcgagccc	ctggagaacg	ccctgaccgt	3060
gttcaccgac	ggcagcagca	acggcaaggc	cgcctacacc	ggccccaagg	agcgcgtat	3120
caagacccccc	taccagagcg	cccagcgc	cgagctggtg	gccgtatca	ccgtgctgca	3180
ggacttcgac	cagccatca	acatcatcag	cgacagcgcc	tacgtggtgc	aggccacccg	3240
cgacgtggag	accgcctga	tcaagtacag	catggacgac	cagctgaacc	agctgtcaa	3300
cctgctgcag	cagaccgtgc	gcaagcgc	cttccccctc	tacatcaccc	acatccgcgc	3360
ccacaccaac	ctgccccggc	ccctgacc	ggcaacag	caggccgacc	tgctggtag	3420
cagcgcctg	atcaaggccc	aggagctgca	cgccctgacc	cacgtgaacg	ccgcccggc	3480
gaagaacaag	ttcgacgtga	ccttggaaagc	ggcaagagc	atcgtcagc	actgcaccc	3540
gtgccagggt	ctgcacctgc	ccacccagga	ggccggcgt	aacccccc	gcctgtgccc	3600
caacgcctg	tggcagatgg	acgtgaccc	cgtgccc	ttcggccg	ttagctacgt	3660
gcacgtgacc	gtggacac	acagccactt	catctggcc	acctgccc	ccggcgag	3720
caccagccac	gtgaagaagc	acctgtcg	ctgcttc	gtatggc	tgcccggaa	3780
gatcaagacc	gacaacggcc	ccggctact	cagcaaggc	ttccagaagt	tcctgagcc	3840
gttgaagatc	agccacacca	ccggcatccc	ctacaac	cagggccagg	ccatcg	3900
gwgccaccaac	cgcacccctg	agacccag	ggtgaagc	aaggagg	gwgacagca	3960
ggagtgcacc	accccccaga	tgcagctgaa	cctggcc	tacacc	acttcc	4020
catctaccgc	aaccagacca	ccaccagc	cgagcag	ccggc	agaagaac	4080
ccccccacag	ggcaagctg	tctggggaa	ggaca	aacaagac	gggagatcg	4140
caaggtgatc	acctggggc	gcggcttc	ctgcgtg	cccggc	accagctg	4200
cgtgtggatc	cccacccggc	acctaagg	ctacaac	ccatcc	acgcca	4260
gagcaccagc	gccgagacc	agaccagcc	gagcag	cacc	aggacg	4320
gaacggcgac	gtgcgcgc	ccgacgagg	ggccat	caggagg	gwgccg	4380
cctgggcacc	accaaggagg	ccgacgccc	gagctaca	atcagcc	agcaca	4440
cgacaccaac	cccccg	acgccc	cagcct	gact	acggcg	4500
gagccccctac	gcctgccc	gcagct	cttaaa	tcagact	gcaagtct	4560
aaagccatgg	atatcgatc	cactac	tagact	tgatc	cgactgt	4620
ttcttagttgc	cagccatct	ttgtt	ctcccc	cctt	ccctgg	4680
tgccactccc	actgtcctt	cctaataaaa	tgaggaa	att	gtctg	4740
gtgtcattct	attctgggg	gtgggg	gcagg	aaaa	gaga	4800
caatagcagg	gggg	gggg	acg	gggg	atggg	4860
agccggcg	ccgg	aaaa	cat	gggg	atggg	4920
cgaaatctcg	tgt	ggc	ggc	gggg	atggg	4980
cgctcagaag	aactcg	taa	ggc	gggg	atggg	5040
gataccgtaa	agc	acg	ggc	gggg	atggg	5100
acgggtagcc	aac	cg	ggc	gggg	atggg	5160
gaatccagaa	aag	cg	ggc	gggg	atggg	5220
cacgacgaga	tcc	cg	ggc	gggg	atggg	5280
cgcgagcccc	tgat	cg	ggc	gggg	atggg	5340
agtacgtgt	cg	ctc	gat	ggc	atggg	5400
aagcgtatgc	agcc	ccg	ccg	ggat	gggg	5460
	ccg	ccg	ccg	ccg	ccg	

Substitute Sequence Listing_USSN 10587032_PP019482.007

gtgagatgac	aggagatct	ccccggcac	ttcgcccaat	agcagccagt	cccttcccgc	5520
ttcagtaca	acgtcgagca	cagctgcga	aggaacgccc	gtcggtggca	gccacgatag	5580
ccgcgtgcc	tcgtcctgca	gttcattcag	ggcacccggac	aggtcggtct	tgacaaaaag	5640
aaccgggcgc	ccctgcgtctg	acagccgaa	cacggcgca	tcaagcagc	cgattgtctg	5700
tttgtccccag	tcatagccga	atagcctctc	cacccaagcg	gcccggagaac	ctgcgtcaa	5760
tccatcttgt	tcaatcatgc	gaaacgatcc	tcatcctgtc	tcttgatcag	atcttgatcc	5820
cctgcgccat	cagatccttg	gcccgaagaa	agccatccag	tttactttgc	agggcttccc	5880
aaccttacca	gagggcgccc	cagctggcaa	ttccggttcg	cttgcgtgtcc	ataaaaaccgc	5940
ccagcttagc	tatcgccatg	taagcccact	gcaagctacc	tgctttctct	ttgcgttgc	6000
gtttccctt	gtccagatag	cccagtagct	gacattcatc	cggggtcagc	accgttctg	6060
cggactggct	ttctacgtgt	tcgcgttcct	tttagcagccc	ttgcgcctgt	agtgcgtgc	6120
gcagcgtgaa	gctaattcat	ggttaaattt	ttgttaaattc	agtcatttt	ttaaccaata	6180
ggccgaaatc	ggccaaaatcc	cttataaattc	aaaagaatag	cccgagatag	ggttgagtgt	6240
tgttccagtt	tggaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaaaggcg	6300
aaaaaccgtc	tatcagggcg	atggccgat	cagcttatgc	ggtgtgaaat	accgcacaga	6360
tgcgttaagga	gaaaataccg	catcaggcgc	tcttccgtt	cctcgctcac	tgactcgctg	6420
cgctcggtcg	ttcggctgcg	gcgagcggta	tcagctact	caaaggcggt	aatacgtta	6480
tccacagaat	caggggataa	cgcaggaaag	aacatgtgag	caaaaggcca	gaaaaaggcc	6540
aggaaccgt	aaaaggccgc	gttgcgtgcg	ttttccata	ggctccggcc	ccctgacgag	6600
catcacaaaa	atcgacgctc	aagttagcagg	tggcgaaacc	cgacaggact	ataaaagatac	6660
caggcgtttc	ccccttggaaag	ctccctcg	cgctctctgt	ttccgcaccc	gcccgttacc	6720
ggataacctgt	ccgcctttt	cccttcggga	agcgtggcgc	ttttctcatag	ctcacgtgt	6780
aggtatctca	gttcgggtgt	ggtcgttcgc	tccaagctgg	gctgtgtgca	cgaacccccc	6840
gttcagcccg	accgctgcgc	cttatccgg	aactatcg	ttgagtc	cccggttaaga	6900
cacgacttat	cgccacttggc	agcagccact	ggtaacagga	ttagcagagc	gaggtatgt	6960
ggcggtgcta	caagttctt	gaagttgtgg	cctaactacg	gctacactag	aaggacagta	7020
tttggtatct	gcgctctgt	gaagccagtt	accttcgaa	aaagagttgg	tagctttga	7080
tccggcaaac	aaaccaccgc	tggtagcgg	ggttttttt	tttgcagca	gcagattacg	7140
cgcagaaaaa	aaggatctca	agaagatcct	ttgatcttt	ctactgaacg	gtgatcccc	7200
ccggaaattgc	g					7211

<210> 57

<211> 318

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 57

atgaacccat	cagagatgca	aagaaaagca	cctccgcgg	gacggagaca	tcgcaatcg	60
gcaccgttga	cttcacaagat	gaacaaaatg	gtgacgtcg	aagaacagat	gaagttgcca	120
tccaccaaga	aggcagagcc	gccaacttgg	gcacaactaa	agaagctgac	gcagttagct	180
acaaaatatac	tagagaacac	aaagggtgaca	caaaccgg	agatgtgt	gcttgcagcc	240
ttgatgattt	tatcaatgg	gtctcgagg	gtacccaaaca	gctccgaaga	gacagcgacc	300
atcgagaacg	ggccatga					318

<210> 58

<211> 321

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified cORF sequence

<400> 58

atgaacccca	gcgagatgca	gwgcaaggcc	cccccccgcc	gccgcgc	ccgcaaccgc	60
gccccccctga	cccacaagat	gaacaagatg	gtgaccagcg	aggagcagat	gaagctgccc	120
agcaccga	aggccgagcc	ccccacccgttgg	gcccagctga	agaagctgac	ccagctggcc	180
accaagtacc	tggagaacac	caagggtgacc	cagacccccc	agagcatgt	gctggccgccc	240
ctgatgatcg	tgagcatgtt	gagcgccggc	gtgcccaca	gcagcgagga	gaccgcccacc	300
atcgagaacg	ggcccgctta	a				321

<210> 59

<211> 435

<212> DNA

Substitute Sequence Listing_USSN 10587032_PP019482.007

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 59

atgaacccat	cggagatgca	aagaaaagca	cctccgcgga	gacggagaca	tcgcaatcga	60
gcaccgttga	ctcacaagat	gaacaaaatg	gtgacgtcag	aagaacagat	gaagttgcc	120
tccaccaga	aggcagagcc	gcacaacttgg	gcacaactaa	agaagctgac	gcagtttagct	180
acaaaatatac	tagagaacac	aaaggtgaca	caaaccagg	agagtatgct	gcttgccg	240
ttgatgattt	tatcaatgg	ggtgtaccca	acagctccg	agagacagcg	accatcgaga	300
acgggcccatt	atgacgatgg	cgggtttgtc	gaaaagaaaa	gggggaaatg	tggggaaaag	360
caagagagat	cagattgtt	ctgtgtctgt	gtagaaaagaa	gtagacatag	gagactccat	420
tttgttctgt	actaa					435

<210> 60

<211> 438

<212> DNA

<213> Artificial Sequence

<220>

<223> Modified PCAP5 sequence

<400> 60

atgaacccca	gcgagatgca	gwgcaaggcc	cccccccgcc	gccgcccggca	ccgcaaccgc	60
ccccccctga	cccacaagat	gaacaagatg	gtgaccagcg	aggagcagat	gaagctgccc	120
agcaccaaga	aggccgagcc	cccccacctgg	gcccagctga	agaagctgac	ccagctggcc	180
accaagtacc	tggagaacac	caaggtgacc	cagaccccg	agagcatgct	gctggccg	240
ctgatgatcg	tgagcatgg	ggtgtacccc	accgccccca	agcgccagcg	ccccagccgc	300
accggccacg	acgacgacgg	cggcttcgtg	gagaagaagc	gcggcaagt	cggcgagaag	360
caggagcgc	gcgactgcta	ctgcgtgtgc	gtggagcgc	gccgcccaccg	ccgcctgcac	420
ttcgtgctgt	acgcttaa					438

<210> 61

<211> 2001

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 61

atggggcaaa	ctaaaagtaa	aattaaaagt	aaatatgcct	cttatctcag	ctttattaaa	60
attctttaa	aaaggggggg	agttaaaagta	tctacaaaaaa	atctaattcaa	gctatttcaa	120
ataatagaac	aattttgccc	atgggttcca	gaacaaggaa	ctttagatct	aaaagattgg	180
aaaagaattt	gtaaggaaact	aaaacaagca	ggttagaagg	gtaatatcat	tccacttaca	240
gtatggaaatg	attgggcccatt	tattaaagca	gcttttagaa	catttcaaac	agaagaagat	300
agcgtttcag	tttctgtatgc	cccttggaa	tgtataatag	attttaatga	aaacacaagg	360
aaaaaatccc	agaaagaaac	ggaaggttt	catttgcgaa	atgttagcaga	gccggtatg	420
gctcagtcaa	cgcaaaatgt	tgactataat	caatttacagg	aggtatata	tcctgaaacg	480
ttaaaattag	aggaaaaagg	tccagaatta	gtggggccat	caagatctaa	accacgaggc	540
acaagtccct	ttccagcagg	tcaggtgcct	gtaacattac	aacctcaaaa	gcaggttaaa	600
gaaaataaga	cccaaccgccc	agtagcctat	caataactgg	ctccggctga	acttcagtat	660
cggccacccc	cagaaaagtca	gtatggat	ccaggaatgc	ccccagcacc	acagggcagg	720
gcgcctatacc	ctcagccgccc	cactaggaga	cttaatccca	cgccaccacc	tagtagacag	780
ggtagtaaat	tacatgaaat	tattgataaa	tcaagaaagg	aaggagatac	tgaggcatgg	840
caattcccag	taacgtttaga	accgtatgca	ccttggagaag	gagcccaaga	gggagagcct	900
cccacagttt	aggccagata	caagtctttt	tcgataaaaa	agctaaaaga	tatgaaagag	960
ggagtaaaac	agtatggacc	caactccct	tatatgagga	cattattaga	ttccattgct	1020
catggacata	gactcattcc	ttatgattgg	gagattctgg	caaaatcg	tctctcaccc	1080
tctcaatttt	tacaattttaa	gacttgggt	attgtatggg	tacaagaaca	ggtccgaaga	1140
aatagggtct	ccaaatccct	agttaacata	gatgcagatc	aactattagg	aataggtcaa	1200
aattggagta	ctattagtca	acaagcatta	atgcaaaaatg	agggcattga	gcaagttaga	1260
gctatctgccc	ttagagcctt	ggaaaaaaatc	caagaccagg	gaagtac	cccctcattt	1320
aatacagtaa	gacaagggtt	aaaagagccc	tatcctgatt	tttgtggcaag	gctccaagat	1380
gttgctcaaa	agtcaatttgc	tgatgaaaaa	gcccgttaagg	tcatagtgga	gttgatggca	1440
tatgaaaacg	ccaaatccct	gtgtcaatca	gccattha	cattaaaagg	aaaggttcct	1500
gcaggatcag	atgttatctc	agaatatgt	aaaggctgt	atgaaatcg	aggagctatg	1560
cataaagcta	tgcttatggc	tcaagcaata	acaggatgt	ttttaggagg	acaagttaga	1620

Substitute Sequence Listing_USSN 10587032_PP019482.007

acatttggaa	aaaaatgtta	taatttgtt	caaattgtc	actaaaaaaaa	gaattgccca	1680
gtcttaaata	aacagaatat	aactattcaa	gcaactacaa	cagtagaga	gccacctgac	1740
ttatgtccaa	gatgtaaaaa	aggaaaacat	tgggctagtc	aatgtcggtc	taaatttgat	1800
aaaaatgggc	aaccattgtc	gggaaacgag	caaaggggcc	agcctcaggc	cccacaacaa	1860
actggggcat	tcccaattca	gccatttgtt	cctcagggtt	ttcagggaca	acaaccccca	1920
ctgtcccaag	tgtttcaggg	aataagccag	ttaccacaat	acaacaattg	tccccgcca	1980
caagcggcag	tgcagcagta	g				2001

<210> 62
<211> 2004
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified gag sequence

<400> 62

atggggcaga	ccaagagcaa	gatcaagagc	aagtacgcca	gctacctgag	cttcatcaag	60
atcctgctga	agcgcggcgg	cgtgaagggt	agcaccaaga	acctgtatcaa	gctgttccag	120
atcatcgagc	agtctgtccc	ctgggttccc	gagcagggca	ccctggacct	gaaggactgg	180
aagcgcacg	gcaaggagat	gaagcaggcc	ggccgcgaagg	gcaacatcat	ccccctgacc	240
gtgtggaaacg	actggggcat	catcaaggcc	gcccctggagc	ccttccagac	cgaggaggac	300
agcgtgagcg	tgagcgcacgc	ccccggcagc	tgcatcatcg	actgcaacga	gaacacccgc	360
aagaagagcc	agaaggagac	cgagggcctg	cactgctgag	acgtggccga	ccccgtgatg	420
gcccagagca	cccagaacgt	ggactacaac	cagctgcagg	aggtgatcta	ccccgagacc	480
ctgaagctgg	agggcaagggg	ccccgagctg	gtggggccca	gcgagagcaa	cccccgccgc	540
accagcccc	tgcccgccgg	ccaggtgccc	gtgaccctgc	agccccagaa	gcaggtgaaag	600
gagaacaaga	cccgcccccc	cgtggctac	cagtactggc	cccccgccga	gctgcagttac	660
cgcccccccc	ccgagagcca	gtacggctac	cccgcatgc	cccccgcccc	ccagggccgc	720
gccccctacc	cccagcccc	cacccgcccgc	ctgaacccca	ccggcccccc	cagccgcccag	780
ggcagcaagc	tgcacgagat	catcgacaag	agccgcaagg	agggcgcacac	cgaggcctgg	840
cagttcccg	tgacccttgg	gcccattgccc	cccggcgagg	gcgcccagga	gggcgagccc	900
cccaccgtgg	aggcccgcta	caagagcttc	agcatcaaga	agctgaaggaa	catgaaggag	960
ggcgtgaagc	agtacggccc	caacagcccc	tacatgcgc	ccctgcttga	cagcatcgcc	1020
cacggccacc	gcctgatccc	ctacgactgg	gagatccctgg	ccaagagcag	cctgagcccc	1080
agccagtttc	tgcagttcaa	gacctgggtgg	atcgacggcg	tgcaggagca	gtgtgcggcgc	1140
aaccgcggcg	ccaaacccccc	cgtgaacatc	gacggccgacc	agctgctggg	catcggccag	1200
aactggagca	ccatcagcca	gcagggccctg	atcgagaacg	aggccatcg	gcaggtgcgc	1260
gccatctgg	tgcgcgcctg	ggagaagatc	caggaccctg	gcagcacctg	ccccagcttc	1320
aacaccgtgc	gccaggggcag	caaggagccc	taccccgact	tcgtggcccg	cctgcaggac	1380
gtggcccaga	agagcatcg	cgacgagaag	gcccgcgaagg	tgatcgttga	gctgtatggcc	1440
tacgagaacg	ccaaaccccg	gtgcccagac	gccatcaagc	ccctgttgggg	caagggtgccc	1500
gccggcagcg	acgtgatcg	cgagttacgt	aaggcctgtcg	acggcatcg	cggcgccatg	1560
cacaaggcca	tgctgatggc	ccaggccatc	accggcgctgg	tgctggggcg	ccaggtgcgc	1620
accttcggcc	gcaagtgtct	caactgcggc	cagatcggtcc	acctgttggaa	gaactgcccc	1680
gtgctgaaca	agcagaacat	caccatccag	gccaccacca	ccggccgcga	cccccccgac	1740
ctgtcccccc	gctgcaagaa	gggcaagcac	tggggccagcc	agtgcgcag	caagttcgac	1800
aagaacggcc	agcccttgcg	cgcaacgcag	cagcgcggcc	agccccaggc	ccccccagcag	1860
accggcgcct	tccccatcca	gcccttcgtt	ccccagggtt	tccaggggca	gcagcccccc	1920
ctgagccagg	tgtttccaggg	catcagccag	ctgccccagt	acaacaactg	cccccccccc	1980
caggccgcgg	tgcagcagggc	ttaa				2004

<210> 63
<211> 1005
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 63

atgtgggcaa	ccattgtcgg	gaaacgagca	aaggggccag	cctcaggccc	cacaacaaac	60
tggggcattc	ccaaattcagc	catttttcc	tcagggtttt	cagggacaac	aaccccccact	120
gtcccaagt	tttcaggaa	taagccagtt	accacaatac	aacaattgtc	ccccgcacaca	180
agcggcagt	cagcgttgc	tttatgtact	atacaagcag	tctctctgt	tccaggggag	240
cccccacaaa	aaaccccccac	agggtataat	ggaccctgtc	ctaaggggac	tgttaggacta	300

Substitute Sequence Listing_USSN 10587032_PP019482.007

atcttggac	gatcaagtct	aaatctaaaa	ggagttcaaa	ttcatactag	tgtggttgat	360
tcagactata	aaggcgaat	tcaattggtt	attagcttt	caattcccttg	gagtgcctgt	420
ccaagagaca	ggattgctca	attattactc	ctgccataca	ttaagggtgg	aaatagtgaa	480
ataaaaagaa	taggagggct	tggaaacact	gatccaacag	gaaaggctgc	atattggca	540
agtcaggtct	cagagaacag	acctgtgtgt	aaggccatta	ttcaaggaaa	acagttgaa	600
gggttggtag	acactggagc	agatgtctct	atcattgctt	taaatcagtg	gccaaaaaaaaat	660
tggccctaaac	aaaaggctgt	tacaggactt	gtcggcatag	gcacagcctc	agaagtgtat	720
caaagtacgg	agattttaca	ttgcttaggg	ccagataatc	aagaaggatc	tgttcagcca	780
atgattactt	caattccctt	taatctgtgg	ggtcgagatt	tattacaaca	atggggtgcg	840
gaaatcacca	tgcccgtcc	atcatatagc	cccacgagtc	aaaaaatcat	gaccaagatg	900
ggatatatac	caggaaaggg	actagggaaa	aatgaagatg	gcattaaaat	tccagttgag	960
gctaaaataa	atcaagaaag	agaaggaata	gggaatccctt	gctag		1005

<210> 64
<211> 1008
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified Prt sequence

<400> 64

atgtgggcca	ccatcgtggg	caagcgcgcc	aaggcccccg	ccagcggccc	caccaccaac	60
tggggcatcc	ccaacagcgc	catctgcagc	agcggcttca	gcgcacccac	caccccccacc	120
gtgcccagcg	tgagcggcaa	caagccgtg	accaccatcc	agcagctgag	ccccgcacc	180
agcggcagcg	ccgcccgtgga	cctgtgcacc	atccagggccg	tgagcctgtct	ccccggcgag	240
cccccccaaga	agaccccccac	cgcgctgtac	ggccccctgc	ccaagggcac	cgtgggcctg	300
atcctgggcc	gcagcagcct	gaacctgaag	ggcgtgcaga	tccacaccag	cgtgggtggac	360
agcgactaca	agggcggagat	ccagctggtg	atcagcagca	gcatccctg	gagcgccagc	420
ccccgcgacc	gcatcgccca	gctgtgtctg	ctgcccatac	tcaagggcg	caacagcgag	480
atcaagcga	tcggcggcct	gggcagcacc	gaccccacccg	gcaaggccgc	ctactggcc	540
agccaggtga	gcgagaacccg	ccccgtgtgc	aaggccatca	tccagggcaa	gcagttcgag	600
ggcctgggtgg	acaccggcgc	cgacgtgagc	atcatcgccc	tgaaccagtg	gcccaagaac	660
tggcccaagc	agaaggccgt	gaccggcctg	gtgggcatcg	gcaccgcccag	c gaggtgtac	720
cagagcaccg	agatcctgca	ctgcctggc	cccgacaaacc	aggagagcac	cgtgcagccc	780
atgatcacca	gcatccccct	gaacctgtgg	ggccgcgacc	tgtcgcagca	gtggggcgcc	840
gagatcacca	tgcccggcccc	cagctacagc	cccaccagcc	agaagatcat	gaccaagatg	900
ggctacatcc	ccggcaaggg	cctggcgaag	aacgaggacg	gcatcaagat	ccccgtggag	960
gccaagatca	accaggagcg	cgagggcatc	ggcaacccct	gcgttaa		1008

<210> 65
<211> 2874
<212> DNA
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 65

atgaataaaat	caagaaaagag	aaggaatagg	gaatccttgc	tagggcggc	cactgttagag	60
cctcctaaac	ccataccatt	aacttggaaa	acagaaaaac	cagtgtgggt	aaatcaagtgg	120
ccgctaccaa	aacaaaaact	ggaggcttta	catttattag	caaatacaca	gttagaaaaag	180
ggtcatattt	agccttcgtt	ctcaccttgg	aattctcttg	tgttgtaat	tcagaagaaa	240
tcaggcaaat	ggcgtatgtt	aactgactta	agggctgtaa	acgccgtaat	tcaacccatg	300
gggcctctcc	aaccgggtt	gccctctccg	gccatgatcc	caaagattg	gcctttaatt	360
ataattgatc	taaaggattt	cttttttacc	atccctctgg	cagacggaga	ttgcgaaaaaa	420
tttgcttta	ctataccagc	cataaataat	aaagaacccag	ccaccagggtt	tcagtggaaa	480
gtgttacctc	agggatgtct	taatagtcca	actatttgc	agacttttgt	aggtcgagct	540
tttcaaccag	tttagagaaaa	gttttccagac	tgtttatatta	ttcattgtat	tgatgtatatt	600
ttatgtctgt	cagaaaacgaa	agataaaatta	attgactgtt	atacattttct	gcaagcagag	660
gttgccatgt	ctggactggc	aatagcatct	gataagatcc	aaacctctac	tccttttcat	720
tattnaggta	tgcagataga	aaatagaaaa	attaagccac	aaaaaataga	aataagaaaa	780
gacacattaa	aaacactaaa	tgattttcaa	aaattactag	gagatattaa	ttggatctgg	840
ccaaactctag	gcattcttac	ttatgccatg	tcaattttgt	tcttatctt	aagaggagac	900
tcagacttaa	atagtaaaag	aatgttaacc	ccagaggcaa	caaagaaaat	taaatttagtg	960
gaagaaaaaa	ttcagtcagc	gcaaaaaat	agaatagatc	ccttagcccc	actccaaactt	1020

Substitute Sequence Listing_USSN 10587032_PP019482.007

ttgatttttgc	ccactgcaca	ttctccaaca	ggcatcatta	ttcaaaatac	tgtatcttgt	1080
gagtggtcat	tccttcctca	cagtacagt	aagacttta	cattgtactt	ggatcaaata	1140
gctacattaa	tcggtcagac	aagattacga	ataataaaaat	tatgtggaa	tgaccaggac	1200
aaaatagttg	tcccttaaac	caaggaacaa	gttagacaag	cctttatcaa	ttctggtca	1260
tggaagatttgc	gtcttgctaa	ttttgtggaa	attattgata	atcattaccc	aaaaacaaag	1320
atcttccagt	tcttaaaatt	gactacttgg	attctaccta	aaattaccag	acgtgaacct	1380
ttagaaaatgc	ctctaacagt	atttactgat	ggttccagca	atgaaaagc	agcttacaca	1440
ggaccgaaag	aacgagtaat	caaaaactcca	tatcaatcg	ctcaaagagc	agagttggtt	1500
gcagtcattaa	cagtgttaca	agattttgac	caacctatca	atattatatc	agattctgca	1560
tatgtgtac	aggctacaag	ggatgtttag	acagctctaa	ttaaatatag	catggatgat	1620
cagttaaacc	agctattcaa	tttattacaa	caaactgtaa	aaaaaagaaa	tttcccattt	1680
tatattacac	atattcgagc	acacactaat	ttaccaggc	ctttgactaa	agcaaatgaa	1740
caagctgact	tactggtatac	atctgcactc	ataaaagcac	aagaacttca	tgctttgact	1800
catgttaaatgc	cagcaggatt	aaaaaacaaa	tttgatgtca	catggaaaca	ggcaaaagat	1860
attgtacaaac	attgcaccca	gtgtcaagtc	ttacaccctgc	ccactcaaga	ggcaggagtt	1920
aatcccagag	gtctgtgtcc	taatgcatta	tggcaaatgg	atgtcacgc	tgtacccatca	1980
tttggaaatgc	tatcatatgt	tcacgttaca	gttgatactt	attcacattt	catatggca	2040
acttgcacaa	caggagaaag	tacttccat	gttaaaaaac	atttattgtc	ttgttttgct	2100
gtatgggag	ttcccgaaaaa	aatcaaaaat	gacaatggac	caggatattg	tagtaaagct	2160
ttcccaaaatgc	tcttaagtca	gtggaaaatt	tcacatataa	cagaattcc	ttataattcc	2220
caaggacagg	ccatagttga	aagaactaat	agaacactca	aaactcaatt	agttaaacaa	2280
aaagaagggg	gagacagtaa	ggagtgtaacc	actcctcaga	tgcactttaa	tctagcactc	2340
tatactttaa	attttttaaa	catttataga	aatcagacta	ctacttctgc	agaacaacat	2400
cttactggta	aaaagaacac	cccacatgaa	ggaaaactaa	tttggtgaa	agataataaa	2460
aataagacat	ggggaaatagg	gaagggtata	acgtggggga	gaggtttgc	ttgtgttca	2520
ccaggagaaa	atcagcttcc	tgtttgata	cccactagac	atttgaagtt	ctacaatgaa	2580
cccatcagag	atgcaaaagaa	aagcacctcc	gcggagacgg	agacatcgca	atcgagcacc	2640
gttgactcac	aatgtgaaca	aaatggtgac	gtcagaagaa	cagatgaagt	tgccatccac	2700
caagaaggca	gagccgccaa	cttggcaca	actaaagaag	ctgacgcagt	tagctacaaa	2760
atatcttagag	aacacaaagg	tgacacaaac	cccagagat	atgtgtcttg	cagccttgat	2820
gattgtatca	atgggtgtaa	gtctccctat	gcctgcagga	gcagctgcag	ctaa	2874

<210> 66
<211> 2877
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified pol sequence

<400> 66						
atgaacaaga	gccgcaagcg	ccgcaaccgc	gagagcctgc	ttggcgccgc	caccgtggag	60
ccccccaaggc	ccatccccct	gaccttggaa	accgagaagc	ccgtgtgggt	gaaccagtgg	120
cccctgccc	agcagaagct	ggaggccc	cacctgctgg	ccaaacgagca	gctggagaag	180
ggccacatcg	agcccagctt	cagccc	aacagcccc	tgttcgttat	ccagaagaag	240
agcggcaagt	ggcgcatgt	gaccgac	cgccgcgtga	acgcccgtat	ccagccatg	300
ggccccctgc	agcccgccct	gcccagcccc	gccatgatcc	ccaaggactg	gcccctgatc	360
atcatcgacc	tgaaggactg	tttcttcacc	atccccctgg	ccgagcagga	ctgcgagaag	420
ttcgccttca	ccatccccgc	catcaacaac	aaggagccc	ccacccgctt	ccagtggaa	480
gtgctgcccc	agggcatgt	gaacagcccc	accatctg	agaccttcgt	ggccgcgccc	540
ctgcagcccg	tgcgcgagaa	gttcagc	tgctacatca	tccactgc	cgacgacatc	600
ctgtgcgc	ccgagaccaa	ggacaagctg	atcgactgt	acaccttcct	gcaggccgag	660
gtggccaacg	ccggccctggc	catcgcc	gacaagatcc	agaccagcac	ccccttccac	720
tacctggca	tgcagatcg	gaaccgca	atcaagcccc	agaagatcg	gatccgc	780
gacacccctga	agacccctgaa	cgacttcc	aagctgtgg	gcgacatca	ctggatccgc	840
cccacccctgg	gcatccccac	ctacgccc	agcaacc	tcagcatc	gcgcggcgac	900
agcgacccctga	acagcaagcg	catgtg	cccggaggc	ccaaaggagat	caagctgg	960
gaggagaaga	tccagagcgc	ccagatcaac	cgcatcg	ccctggccc	cctgcagctg	1020
ctgatcttcg	ccacccgccc	cagccccacc	ggc	tccagaacac	cgacccctgg	1080
gagtggagct	tccctggccca	cagcaccgtg	aagacttca	ccctgtac	ggaccagatc	1140
gccacccctga	tcggccagac	ccgcctgc	atcatca	tgtgcggca	cgaccccgac	1200
aagatctgttgc	tgccccctgac	caaggagc	gtgcgc	ccttcatca	cagcggcgcc	1260
tggaaatcg	gcctggccaa	ttcgtggc	atcatcg	accactaccc	caagaccaag	1320

Substitute Sequence Listing_USSN 10587032_PP019482.007

atcttccagt	tcctgaagct	gaccacctgg	atcctgcccc	agatcaccgg	ccgcgagccc	1380
ctggagaacg	ccctgaccgt	gttcaccgac	ggcagcagca	acggcaaggc	cgcctacacc	1440
ggccccaagg	agcgctgtat	caagaccccc	taccagagcg	cccagcgcgc	cgagctggtg	1500
gccgtatca	ccgtgctgca	ggacttcgac	cagccccatca	acatcatcg	cgacagcgcc	1560
tacgtgggtc	aggccacccg	cgacgtggag	accgcccgt	tcaagtatcg	catggacgac	1620
cagctgaacc	agctgttcaa	cctgctgcag	cagaccgtgc	gcaagcgaa	cttcccttc	1680
tacatcaccc	acatccgcgc	ccacaccaac	ctgcccggcc	ccctgaccaa	ggccaacgag	1740
caggccgacc	tgctggtag	cagccccgt	atcaaggccc	aggagctgca	cgcctgtacc	1800
cacgtgaacg	ccgccccgt	gaagaacaag	ttcgcacgtga	cctggaagca	ggccaaggac	1860
atcgtgcagc	actgcaccc	gtgccagggt	ctgcacctgc	ccacccagga	ggccggcgtg	1920
aaccccccgcg	gcctgtgccc	caacgcccc	tggcagatgg	acgtgaccca	cgtgcccagc	1980
ttcggccgccc	tgagctacgt	gcacgtgacc	gtggacacct	acagccactt	catctggcc	2040
acctgcccaga	ccggcgagag	caccagccac	gtgaagaagc	acctgcttag	ctgcttcgccc	2100
gtgatgggcg	tgcccggagaa	gatcaagacc	gacaacggcc	ccggctactg	cagcaaggcc	2160
ttccagaagt	tcctgagcca	gtggaagatc	agccacacca	ccggcatccc	ctacaacagc	2220
cagggccagg	ccatcggtgg	gcbcaccaac	cgccacccgt	agacccagct	ggtgaagcag	2280
aaggagggcg	gcbcacagcaa	ggagtgcacc	accccccaga	tgcagctgaa	cctggccctg	2340
tacacccgt	acttcctgaa	catctaccgc	aaccagacca	ccaccagcgc	cgagcagcac	2400
ctgaccggc	agaagaacag	ccccacacgag	ggcaagctga	tctgggtggaa	ggacaacaag	2460
aacaagacct	gggagatgg	caagggtgatc	acctggggcc	gcccgttgc	ctgcgtgagc	2520
cccggcgaga	accagctgccc	cgtgtggatc	cccacccccc	acctaaggat	ctacaacgag	2580
cccaccccg	acgccaagaaa	gaggcaccgc	gccgagaccc	agaccagcca	gaggcaccgc	2640
gtggacagcc	aggacgagca	gaacggcgac	gtgcggcgc	ccgacgagg	ggccatccac	2700
caggagggcc	gcccggccaa	cctgggcacc	accaaggagg	ccgacgcccgt	gagctacaag	2760
atcagccgcg	agcacaagg	cgacaccaac	ccccgcgagt	acggccctg	cagcctggac	2820
gactgcatca	acggcgccaa	gagccctac	gcctggcgc	gcagctgca	cgcttaa	2877

<210> 67

<211> 106

<212> PRT

<213> Artificial Sequence

<220>

<223> Manipulated cORF

<400> 67

Met	Asn	Pro	Ser	Glu	Met	Gln	Arg	Lys	Ala	Pro	Pro	Arg	Arg	Arg	Arg
1					5				10					15	

His	Arg	Asn	Arg	Ala	Pro	Leu	Thr	His	Lys	Met	Asn	Lys	Met	Val	Thr
								20		25			30		

Ser	Glu	Glu	Gln	Met	Lys	Leu	Pro	Ser	Thr	Lys	Lys	Ala	Glu	Pro	Pro
								35	40			45			

Thr	Trp	Ala	Gln	Leu	Lys	Lys	Leu	Thr	Gln	Leu	Ala	Thr	Lys	Tyr	Leu
								50	55			60			

Glu	Asn	Thr	Lys	Val	Thr	Gln	Thr	Pro	Glu	Ser	Met	Leu	Leu	Ala	Ala
								65	70		75		80		

Leu	Met	Ile	Val	Ser	Met	Val	Ser	Ala	Gly	Val	Pro	Asn	Ser	Ser	Glu
								85	90			95			

Glu	Thr	Ala	Thr	Ile	Glu	Asn	Gly	Pro	Ala
				100				105	

<210> 68

<211> 145

<212> PRT

<213> Artificial Sequence

<220>

Substitute Sequence Listing_USSN 10587032_PP019482.007
<223> Manipulated PCAP5

<400> 68
Met Asn Pro Ser Glu Met Gln Arg Lys Ala Pro Pro Arg Arg Arg Arg
1 5 10 15

His Arg Asn Arg Ala Pro Leu Thr His Lys Met Asn Lys Met Val Thr
20 25 30

Ser Glu Glu Gln Met Lys Leu Pro Ser Thr Lys Lys Ala Glu Pro Pro
35 40 45

Thr Trp Ala Gln Leu Lys Lys Leu Thr Gln Leu Ala Thr Lys Tyr Leu
50 55 60

Glu Asn Thr Lys Val Thr Gln Thr Pro Glu Ser Met Leu Leu Ala Ala
65 70 75 80

Leu Met Ile Val Ser Met Val Val Tyr Pro Thr Ala Pro Lys Arg Gln
85 90 95

Arg Pro Ser Arg Thr Gly His Asp Asp Gly Gly Phe Val Glu Lys
100 105 110

Lys Arg Gly Lys Cys Gly Glu Lys Gln Glu Arg Ser Asp Cys Tyr Cys
115 120 125

Val Cys Val Glu Arg Ser Arg His Arg Arg Leu His Phe Val Leu Tyr
130 135 140

Ala
145

<210> 69
<211> 666
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 69
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu
1 5 10 15

Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr
20 25 30

Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp
35 40 45

Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly
50 55 60

Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr
65 70 75 80

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
85 90 95

Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
100 105 110

Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
115 120 125

Substitute Sequence Listing_USSN 10587032_PP019482.007

Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
130 135 140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
145 150 155 160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
165 170 175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
180 185 190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
195 200 205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
210 215 220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
225 230 235 240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
245 250 255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
260 265 270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
275 280 285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
290 295 300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
305 310 315 320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
325 330 335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
340 345 350

Leu Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
355 360 365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
370 375 380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln
385 390 395 400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
405 410 415

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Glu Lys Ile Gln Asp
420 425 430

Pro Gly Ser Thr Cys Pro Ser Phe Asn Thr Val Arg Gln Gly Ser Lys
435 440 445

Glu Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Val Ala Gln Lys
450 455 460

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ser Ile Ala Asp Glu Lys Ala Arg Lys Val Ile Val Glu Leu Met Ala			
465	470	475	480
Tyr Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys			
485	490	495	
Gly Lys Val Pro Ala Gly Ser Asp Val Ile Ser Glu Tyr Val Lys Ala			
500	505	510	
Cys Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln			
515	520	525	
Ala Ile Thr Gly Val Val Leu Gly Gly Gln Val Arg Thr Phe Gly Arg			
530	535	540	
Lys Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Lys Asn Cys Pro			
545	550	555	560
Val Leu Asn Lys Gln Asn Ile Thr Ile Gln Ala Thr Thr Thr Gly Arg			
565	570	575	
Glu Pro Pro Asp Leu Cys Pro Arg Cys Lys Lys Gly Lys His Trp Ala			
580	585	590	
Ser Gln Cys Arg Ser Lys Phe Asp Lys Asn Gly Gln Pro Leu Ser Gly			
595	600	605	
Asn Glu Gln Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe			
610	615	620	
Pro Ile Gln Pro Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Pro			
625	630	635	640
Leu Ser Gln Val Phe Gln Gly Ile Ser Gln Leu Pro Gln Tyr Asn Asn			
645	650	655	
Cys Pro Pro Pro Gln Ala Ala Val Gln Gln			
660	665		
<210> 70			
<211> 667			
<212> PRT			
<213> Artificial Sequence			
<220>			
<223> Manipulated Gag			
<400> 70			
Met Gly Gln Thr Lys Ser Lys Ile Lys Ser Lys Tyr Ala Ser Tyr Leu			
1	5	10	15
Ser Phe Ile Lys Ile Leu Leu Lys Arg Gly Gly Val Lys Val Ser Thr			
20	25	30	
Lys Asn Leu Ile Lys Leu Phe Gln Ile Ile Glu Gln Phe Cys Pro Trp			
35	40	45	
Phe Pro Glu Gln Gly Thr Leu Asp Leu Lys Asp Trp Lys Arg Ile Gly			
50	55	60	
Lys Glu Leu Lys Gln Ala Gly Arg Lys Gly Asn Ile Ile Pro Leu Thr			
65	70	75	80

Substitute Sequence Listing_USSN 10587032_PP019482.007

Val Trp Asn Asp Trp Ala Ile Ile Lys Ala Ala Leu Glu Pro Phe Gln
 85 90 95

Thr Glu Glu Asp Ser Val Ser Val Ser Asp Ala Pro Gly Ser Cys Ile
 100 105 110

Ile Asp Cys Asn Glu Asn Thr Arg Lys Lys Ser Gln Lys Glu Thr Glu
 115 120 125

Gly Leu His Cys Glu Tyr Val Ala Glu Pro Val Met Ala Gln Ser Thr
 130 135 140

Gln Asn Val Asp Tyr Asn Gln Leu Gln Glu Val Ile Tyr Pro Glu Thr
 145 150 155 160

Leu Lys Leu Glu Gly Lys Gly Pro Glu Leu Val Gly Pro Ser Glu Ser
 165 170 175

Lys Pro Arg Gly Thr Ser Pro Leu Pro Ala Gly Gln Val Pro Val Thr
 180 185 190

Leu Gln Pro Gln Lys Gln Val Lys Glu Asn Lys Thr Gln Pro Pro Val
 195 200 205

Ala Tyr Gln Tyr Trp Pro Pro Ala Glu Leu Gln Tyr Arg Pro Pro Pro
 210 215 220

Glu Ser Gln Tyr Gly Tyr Pro Gly Met Pro Pro Ala Pro Gln Gly Arg
 225 230 235 240

Ala Pro Tyr Pro Gln Pro Pro Thr Arg Arg Leu Asn Pro Thr Ala Pro
 245 250 255

Pro Ser Arg Gln Gly Ser Lys Leu His Glu Ile Ile Asp Lys Ser Arg
 260 265 270

Lys Glu Gly Asp Thr Glu Ala Trp Gln Phe Pro Val Thr Leu Glu Pro
 275 280 285

Met Pro Pro Gly Glu Gly Ala Gln Glu Gly Glu Pro Pro Thr Val Glu
 290 295 300

Ala Arg Tyr Lys Ser Phe Ser Ile Lys Lys Leu Lys Asp Met Lys Glu
 305 310 315 320

Gly Val Lys Gln Tyr Gly Pro Asn Ser Pro Tyr Met Arg Thr Leu Leu
 325 330 335

Asp Ser Ile Ala His Gly His Arg Leu Ile Pro Tyr Asp Trp Glu Ile
 340 345 350

Leu Ala Lys Ser Ser Leu Ser Pro Ser Gln Phe Leu Gln Phe Lys Thr
 355 360 365

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Arg Asn Arg Ala Ala
 370 375 380

Asn Pro Pro Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Ile Gly Gln
 385 390 395 400

Asn Trp Ser Thr Ile Ser Gln Gln Ala Leu Met Gln Asn Glu Ala Ile
 405 410 415

Substitute Sequence Listing_USSN 10587032_PP019482.007

Glu	Gln	Val	Arg	Ala	Ile	Cys	Leu	Arg	Ala	Trp	Glu	Lys	Ile	Gln	Asp
420							425					430			
Pro	Gly	Ser	Thr	Cys	Pro	Ser	Phe	Asn	Thr	Val	Arg	Gln	Gly	Ser	Lys
435							440					445			
Glu	Pro	Tyr	Pro	Asp	Phe	Val	Ala	Arg	Leu	Gln	Asp	Val	Ala	Gln	Lys
450							455					460			
Ser	Ile	Ala	Asp	Glu	Lys	Ala	Arg	Lys	Val	Ile	Val	Glu	Leu	Met	Ala
465							470					475			480
Tyr	Glu	Asn	Ala	Asn	Pro	Glu	Cys	Gln	Ser	Ala	Ile	Lys	Pro	Leu	Lys
485							490						495		
Gly	Lys	Val	Pro	Ala	Gly	Ser	Asp	Val	Ile	Ser	Glu	Tyr	Val	Lys	Ala
500							505						510		
Cys	Asp	Gly	Ile	Gly	Gly	Ala	Met	His	Lys	Ala	Met	Leu	Met	Ala	Gln
515							520						525		
Ala	Ile	Thr	Gly	Val	Val	Leu	Gly	Gly	Gln	Val	Arg	Thr	Phe	Gly	Arg
530							535						540		
Lys	Cys	Tyr	Asn	Cys	Gly	Gln	Ile	Gly	His	Leu	Lys	Lys	Asn	Cys	Pro
545							550						555		
Val	Leu	Asn	Lys	Gln	Asn	Ile	Thr	Ile	Gln	Ala	Thr	Thr	Thr	Gly	Arg
565							570						575		
Glu	Pro	Pro	Asp	Leu	Cys	Pro	Arg	Cys	Lys	Lys	Gly	Lys	His	Trp	Ala
580							585						590		
Ser	Gln	Cys	Arg	Ser	Lys	Phe	Asp	Lys	Asn	Gly	Gln	Pro	Leu	Ser	Gly
595							600						605		
Asn	Glu	Gln	Arg	Gly	Gln	Pro	Gln	Ala	Pro	Gln	Gln	Thr	Gly	Ala	Phe
610							615						620		
Pro	Ile	Gln	Pro	Phe	Val	Pro	Gln	Gly	Phe	Gln	Gly	Gln	Gln	Pro	Pro
625							630						635		
Leu	Ser	Gln	Val	Phe	Gln	Gly	Ile	Ser	Gln	Leu	Pro	Gln	Tyr	Asn	Asn
645							650						655		
Cys	Pro	Pro	Pro	Gln	Ala	Ala	Val	Gln	Gln	Ala					
660							665								
<210>	71														
<211>	334														
<212>	PRT														
<213>	Human endogenous retrovirus, K family (HERV-K)														
<400>	71														
Met	Trp	Ala	Thr	Ile	Val	Gly	Lys	Arg	Ala	Lys	Gly	Pro	Ala	Ser	Gly
1							5					10			15
Pro	Thr	Thr	Asn	Trp	Gly	Ile	Pro	Asn	Ser	Ala	Ile	Cys	Ser	Ser	Gly
20							25						30		
Phe	Ser	Gly	Thr	Thr	Pro	Thr	Val	Pro	Ser	Val	Ser	Gly	Asn	Lys	
35							40						45		

Substitute Sequence Listing_USSN 10587032_PP019482.007

Pro	Val	Thr	Thr	Ile	Gln	Gln	Leu	Ser	Pro	Ala	Thr	Ser	Gly	Ser	Ala
	50				55				60						
Ala	Val	Asp	Leu	Cys	Thr	Ile	Gln	Ala	Val	Ser	Leu	Leu	Pro	Gly	Glu
	65				70			75		80					
Pro	Pro	Gln	Lys	Thr	Pro	Thr	Gly	Val	Tyr	Gly	Pro	Leu	Pro	Lys	Gly
		85				90			95						
Thr	Val	Gly	Leu	Ile	Leu	Gly	Arg	Ser	Ser	Leu	Asn	Leu	Lys	Gly	Val
		100				105				110					
Gln	Ile	His	Thr	Ser	Val	Val	Asp	Ser	Asp	Tyr	Lys	Gly	Glu	Ile	Gln
		115				120				125					
Leu	Val	Ile	Ser	Ser	Ser	Ile	Pro	Trp	Ser	Ala	Ser	Pro	Arg	Asp	Arg
		130				135			140						
Ile	Ala	Gln	Leu	Leu	Leu	Leu	Pro	Tyr	Ile	Lys	Gly	Gly	Asn	Ser	Glu
		145				150			155				160		
Ile	Lys	Arg	Ile	Gly	Gly	Leu	Gly	Ser	Thr	Asp	Pro	Thr	Gly	Lys	Ala
		165				170			175						
Ala	Tyr	Trp	Ala	Ser	Gln	Val	Ser	Glu	Asn	Arg	Pro	Val	Cys	Lys	Ala
		180				185			190						
Ile	Ile	Gln	Gly	Lys	Gln	Phe	Glu	Gly	Leu	Val	Asp	Thr	Gly	Ala	Asp
		195				200			205						
val	Ser	Ile	Ile	Ala	Leu	Asn	Gln	Trp	Pro	Lys	Asn	Trp	Pro	Lys	Gln
		210				215			220						
Lys	Ala	Val	Thr	Gly	Leu	Val	Gly	Ile	Gly	Thr	Ala	Ser	Glu	Val	Tyr
		225				230			235			240			
Gln	Ser	Thr	Glu	Ile	Leu	His	Cys	Leu	Gly	Pro	Asp	Asn	Gln	Glu	Ser
		245					250				255				
Thr	Val	Gln	Pro	Met	Ile	Thr	Ser	Ile	Pro	Leu	Asn	Leu	Trp	Gly	Arg
		260				265			270						
Asp	Leu	Leu	Gln	Gln	Trp	Gly	Ala	Glu	Ile	Thr	Met	Pro	Ala	Pro	Ser
		275				280			285						
Tyr	Ser	Pro	Thr	Ser	Gln	Lys	Ile	Met	Thr	Lys	Met	Gly	Tyr	Ile	Pro
		290				295			300						
Gly	Lys	Gly	Leu	Gly	Lys	Asn	Glu	Asp	Gly	Ile	Lys	Ile	Pro	Val	Glu
		305				310			315			320			
Ala	Lys	Ile	Asn	Gln	Glu	Arg	Glu	Gly	Ile	Gly	Asn	Pro	Cys		
			325				330								

<210> 72
<211> 335
<212> PRT
<213> Artificial Sequence

<220>
<223> Manipulated Prt

<400> 72

Substitute Sequence Listing_USSN 10587032_PP019482.007

Met Trp Ala Thr Ile Val Gly Lys Arg Ala Lys Gly Pro Ala Ser Gly
1 5 10 15

Pro Thr Thr Asn Trp Gly Ile Pro Asn Ser Ala Ile Cys Ser Ser Gly
20 25 30

Phe Ser Gly Thr Thr Pro Thr Val Pro Ser Val Ser Gly Asn Lys
35 40 45

Pro Val Thr Thr Ile Gln Gln Leu Ser Pro Ala Thr Ser Gly Ser Ala
50 55 60

Ala Val Asp Leu Cys Thr Ile Gln Ala Val Ser Leu Leu Pro Gly Glu
65 70 75 80

Pro Pro Gln Lys Thr Pro Thr Gly Val Tyr Gly Pro Leu Pro Lys Gly
85 90 95

Thr Val Gly Leu Ile Leu Gly Arg Ser Ser Leu Asn Leu Lys Gly Val
100 105 110

Gln Ile His Thr Ser Val Val Asp Ser Asp Tyr Lys Gly Glu Ile Gln
115 120 125

Leu Val Ile Ser Ser Ser Ile Pro Trp Ser Ala Ser Pro Arg Asp Arg
130 135 140

Ile Ala Gln Leu Leu Leu Leu Pro Tyr Ile Lys Gly Gly Asn Ser Glu
145 150 155 160

Ile Lys Arg Ile Gly Gly Leu Gly Ser Thr Asp Pro Thr Gly Lys Ala
165 170 175

Ala Tyr Trp Ala Ser Gln Val Ser Glu Asn Arg Pro Val Cys Lys Ala
180 185 190

Ile Ile Gln Gly Lys Gln Phe Glu Gly Leu Val Asp Thr Gly Ala Asp
195 200 205

Val Ser Ile Ile Ala Leu Asn Gln Trp Pro Lys Asn Trp Pro Lys Gln
210 215 220

Lys Ala Val Thr Gly Leu Val Gly Ile Gly Thr Ala Ser Glu Val Tyr
225 230 235 240

Gln Ser Thr Glu Ile Leu His Cys Leu Gly Pro Asp Asn Gln Glu Ser
245 250 255

Thr Val Gln Pro Met Ile Thr Ser Ile Pro Leu Asn Leu Trp Gly Arg
260 265 270

Asp Leu Leu Gln Gln Trp Gly Ala Glu Ile Thr Met Pro Ala Pro Ser
275 280 285

Tyr Ser Pro Thr Ser Gln Lys Ile Met Thr Lys Met Gly Tyr Ile Pro
290 295 300

Gly Lys Gly Leu Gly Lys Asn Glu Asp Gly Ile Lys Ile Pro Val Glu
305 310 315 320

Ala Lys Ile Asn Gln Glu Arg Glu Gly Ile Gly Asn Pro Cys Ala
325 330 335

Substitute Sequence Listing_USSN 10587032_PP019482.007

<210> 73
<211> 957
<212> PRT
<213> Human endogenous retrovirus, K family (HERV-K)

<400> 73
Met Asn Lys Ser Arg Lys Arg Arg Asn Arg Glu Ser Leu Leu Gly Ala
1 5 10 15
Ala Thr Val Glu Pro Pro Lys Pro Ile Pro Leu Thr Trp Lys Thr Glu
20 25 30
Lys Pro Val Trp Val Asn Gln Trp Pro Leu Pro Lys Gln Lys Leu Glu
35 40 45
Ala Leu His Leu Leu Ala Asn Glu Gln Leu Glu Lys Gly His Ile Glu
50 55 60
Pro Ser Phe Ser Pro Trp Asn Ser Pro Val Phe Val Ile Gln Lys Lys
65 70 75 80
Ser Gly Lys Trp Arg Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val
85 90 95
Ile Gln Pro Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met
100 105 110
Ile Pro Lys Asp Trp Pro Leu Ile Ile Asp Leu Lys Asp Cys Phe
115 120 125
Phe Thr Ile Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr
130 135 140
Ile Pro Ala Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys
145 150 155 160
Val Leu Pro Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe
165 170 175
Val Gly Arg Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr
180 185 190
Ile Ile His Cys Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp
195 200 205
Lys Leu Ile Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala
210 215 220
Gly Leu Ala Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His
225 230 235 240
Tyr Leu Gly Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile
245 250 255
Glu Ile Arg Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu
260 265 270
Leu Gly Asp Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr
275 280 285
Ala Met Ser Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn
290 295 300

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ser Lys Arg Met Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val
305 310 315 320

Glu Glu Lys Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala
325 330 335

Pro Leu Gln Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile
340 345 350

Ile Ile Gln Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser
355 360 365

Thr Val Lys Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile
370 375 380

Gly Gln Thr Arg Leu Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp
385 390 395 400

Lys Ile Val Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile
405 410 415

Asn Ser Gly Ala Trp Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile
420 425 430

Asp Asn His Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr
435 440 445

Thr Trp Ile Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala
450 455 460

Leu Thr Val Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr
465 470 475 480

Gly Pro Lys Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg
485 490 495

Ala Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro
500 505 510

Ile Asn Ile Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp
515 520 525

Val Glu Thr Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln
530 535 540

Leu Phe Asn Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe
545 550 555 560

Tyr Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr
565 570 575

Lys Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys
580 585 590

Ala Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys
595 600 605

Asn Lys Phe Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His
610 615 620

Cys Thr Gln Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val
625 630 635 640

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asn	Pro	Arg	Gly	Leu	Cys	Pro	Asn	Ala	Leu	Trp	Gln	Met	Asp	Val	Thr
			645			650						655			
His	Val	Pro	Ser	Phe	Gly	Arg	Leu	Ser	Tyr	Val	His	Val	Thr	Val	Asp
	660			665						670					
Thr	Tyr	Ser	His	Phe	Ile	Trp	Ala	Thr	Cys	Gln	Thr	Gly	Glu	Ser	Thr
	675			680						685					
Ser	His	Val	Lys	Lys	His	Leu	Leu	Ser	Cys	Phe	Ala	Val	Met	Gly	Val
	690			695						700					
Pro	Glu	Lys	Ile	Lys	Thr	Asp	Asn	Gly	Pro	Gly	Tyr	Cys	Ser	Lys	Ala
	705			710				715						720	
Phe	Gln	Lys	Phe	Leu	Ser	Gln	Trp	Lys	Ile	Ser	His	Thr	Thr	Gly	Ile
	725					730						735			
Pro	Tyr	Asn	Ser	Gln	Gly	Gln	Ala	Ile	Val	Glu	Arg	Thr	Asn	Arg	Thr
	740					745				750					
Leu	Lys	Thr	Gln	Leu	Val	Lys	Gln	Lys	Glu	Gly	Gly	Asp	Ser	Lys	Glu
	755					760				765					
Cys	Thr	Thr	Pro	Gln	Met	Gln	Leu	Asn	Leu	Ala	Leu	Tyr	Thr	Leu	Asn
	770					775				780					
Phe	Leu	Asn	Ile	Tyr	Arg	Asn	Gln	Thr	Thr	Thr	Ser	Ala	Glu	Gln	His
	785					790			795					800	
Leu	Thr	Gly	Lys	Lys	Asn	Ser	Pro	His	Glu	Gly	Lys	Leu	Ile	Trp	Trp
	805						810					815			
Lys	Asp	Asn	Lys	Asn	Lys	Thr	Trp	Glu	Ile	Gly	Lys	Val	Ile	Thr	Trp
	820					825						830			
Gly	Arg	Gly	Phe	Ala	Cys	Val	Ser	Pro	Gly	Glu	Asn	Gln	Leu	Pro	Val
	835					840				845					
Trp	Ile	Pro	Thr	Arg	His	Leu	Lys	Phe	Tyr	Asn	Glu	Pro	Ile	Arg	Asp
	850					855				860					
Ala	Lys	Lys	Ser	Thr	Ser	Ala	Glu	Thr	Glu	Thr	Ser	Gln	Ser	Ser	Thr
	865					870			875					880	
Val	Asp	Ser	Gln	Asp	Glu	Gln	Asn	Gly	Asp	Val	Arg	Arg	Thr	Asp	Glu
	885						890				895				
Val	Ala	Ile	His	Gln	Glu	Gly	Arg	Ala	Ala	Asn	Leu	Gly	Thr	Thr	Lys
	900					905					910				
Glu	Ala	Asp	Ala	Val	Ser	Tyr	Lys	Ile	Ser	Arg	Glu	His	Lys	Gly	Asp
	915					920				925					
Thr	Asn	Pro	Arg	Glu	Tyr	Ala	Ala	Cys	Ser	Leu	Asp	Asp	Cys	Ile	Asn
	930					935				940					
Gly	Gly	Lys	Ser	Pro	Tyr	Ala	Cys	Arg	Ser	Ser	Cys	Ser			
	945					950				955					

<210> 74
<211> 958
<212> PRT

Substitute Sequence Listing_USSN 10587032_PP019482.007
<213> Artificial Sequence

<220>

<223> Manipulated Pol

<400> 74

Met Asn Lys Ser Arg Lys Arg Arg Asn Arg Glu Ser Leu Leu Gly Ala
1 5 10 15

Ala Thr Val Glu Pro Pro Lys Pro Ile Pro Leu Thr Trp Lys Thr Glu
20 25 30

Lys Pro Val Trp Val Asn Gln Trp Pro Leu Pro Lys Gln Lys Leu Glu
35 40 45

Ala Leu His Leu Leu Ala Asn Glu Gln Leu Glu Lys Gly His Ile Glu
50 55 60

Pro Ser Phe Ser Pro Trp Asn Ser Pro Val Phe Val Ile Gln Lys Lys
65 70 75 80

Ser Gly Lys Trp Arg Met Leu Thr Asp Leu Arg Ala Val Asn Ala Val
85 90 95

Ile Gln Pro Met Gly Pro Leu Gln Pro Gly Leu Pro Ser Pro Ala Met
100 105 110

Ile Pro Lys Asp Trp Pro Leu Ile Ile Asp Leu Lys Asp Cys Phe
115 120 125

Phe Thr Ile Pro Leu Ala Glu Gln Asp Cys Glu Lys Phe Ala Phe Thr
130 135 140

Ile Pro Ala Ile Asn Asn Lys Glu Pro Ala Thr Arg Phe Gln Trp Lys
145 150 155 160

Val Leu Pro Gln Gly Met Leu Asn Ser Pro Thr Ile Cys Gln Thr Phe
165 170 175

Val Gly Arg Ala Leu Gln Pro Val Arg Glu Lys Phe Ser Asp Cys Tyr
180 185 190

Ile Ile His Cys Ile Asp Asp Ile Leu Cys Ala Ala Glu Thr Lys Asp
195 200 205

Lys Leu Ile Asp Cys Tyr Thr Phe Leu Gln Ala Glu Val Ala Asn Ala
210 215 220

Gly Leu Ala Ile Ala Ser Asp Lys Ile Gln Thr Ser Thr Pro Phe His
225 230 235 240

Tyr Leu Gly Met Gln Ile Glu Asn Arg Lys Ile Lys Pro Gln Lys Ile
245 250 255

Glu Ile Arg Lys Asp Thr Leu Lys Thr Leu Asn Asp Phe Gln Lys Leu
260 265 270

Leu Gly Asp Ile Asn Trp Ile Arg Pro Thr Leu Gly Ile Pro Thr Tyr
275 280 285

Ala Met Ser Asn Leu Phe Ser Ile Leu Arg Gly Asp Ser Asp Leu Asn
290 295 300

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ser Lys Arg Met Leu Thr Pro Glu Ala Thr Lys Glu Ile Lys Leu Val
305 310 315 320

Glu Glu Lys Ile Gln Ser Ala Gln Ile Asn Arg Ile Asp Pro Leu Ala
325 330 335

Pro Leu Gln Leu Leu Ile Phe Ala Thr Ala His Ser Pro Thr Gly Ile
340 345 350

Ile Ile Gln Asn Thr Asp Leu Val Glu Trp Ser Phe Leu Pro His Ser
355 360 365

Thr Val Lys Thr Phe Thr Leu Tyr Leu Asp Gln Ile Ala Thr Leu Ile
370 375 380

Gly Gln Thr Arg Leu Arg Ile Ile Lys Leu Cys Gly Asn Asp Pro Asp
385 390 395 400

Lys Ile Val Val Pro Leu Thr Lys Glu Gln Val Arg Gln Ala Phe Ile
405 410 415

Asn Ser Gly Ala Trp Lys Ile Gly Leu Ala Asn Phe Val Gly Ile Ile
420 425 430

Asp Asn His Tyr Pro Lys Thr Lys Ile Phe Gln Phe Leu Lys Leu Thr
435 440 445

Thr Trp Ile Leu Pro Lys Ile Thr Arg Arg Glu Pro Leu Glu Asn Ala
450 455 460

Leu Thr Val Phe Thr Asp Gly Ser Ser Asn Gly Lys Ala Ala Tyr Thr
465 470 475 480

Gly Pro Lys Glu Arg Val Ile Lys Thr Pro Tyr Gln Ser Ala Gln Arg
485 490 495

Ala Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asp Gln Pro
500 505 510

Ile Asn Ile Ile Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Arg Asp
515 520 525

Val Glu Thr Ala Leu Ile Lys Tyr Ser Met Asp Asp Gln Leu Asn Gln
530 535 540

Leu Phe Asn Leu Leu Gln Gln Thr Val Arg Lys Arg Asn Phe Pro Phe
545 550 555 560

Tyr Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr
565 570 575

Lys Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Leu Ile Lys
580 585 590

Ala Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ala Gly Leu Lys
595 600 605

Asn Lys Phe Asp Val Thr Trp Lys Gln Ala Lys Asp Ile Val Gln His
610 615 620

Cys Thr Gln Cys Gln Val Leu His Leu Pro Thr Gln Glu Ala Gly Val
625 630 635 640

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asn	Pro	Arg	Gly	Leu	Cys	Pro	Asn	Ala	Leu	Trp	Gln	Met	Asp	Val	Thr
				645					650			655			
His	Val	Pro	Ser	Phe	Gly	Arg	Leu	Ser	Tyr	Val	His	Val	Thr	Val	Asp
	660				665				670						
Thr	Tyr	Ser	His	Phe	Ile	Trp	Ala	Thr	Cys	Gln	Thr	Gly	Glu	Ser	Thr
	675				680				685						
Ser	His	Val	Lys	Lys	His	Leu	Leu	Ser	Cys	Phe	Ala	Val	Met	Gly	Val
	690				695				700						
Pro	Glu	Lys	Ile	Lys	Thr	Asp	Asn	Gly	Pro	Gly	Tyr	Cys	Ser	Lys	Ala
	705				710				715			720			
Phe	Gln	Lys	Phe	Leu	Ser	Gln	Trp	Lys	Ile	Ser	His	Thr	Thr	Gly	Ile
	725					730				735					
Pro	Tyr	Asn	Ser	Gln	Gly	Gln	Ala	Ile	Val	Glu	Arg	Thr	Asn	Arg	Thr
	740					745				750					
Leu	Lys	Thr	Gln	Leu	Val	Lys	Gln	Lys	Glu	Gly	Gly	Asp	Ser	Lys	Glu
	755					760				765					
Cys	Thr	Thr	Pro	Gln	Met	Gln	Leu	Asn	Leu	Ala	Leu	Tyr	Thr	Leu	Asn
	770					775				780					
Phe	Leu	Asn	Ile	Tyr	Arg	Asn	Gln	Thr	Thr	Thr	Ser	Ala	Glu	Gln	His
	785					790			795			800			
Leu	Thr	Gly	Lys	Lys	Asn	Ser	Pro	His	Glu	Gly	Lys	Leu	Ile	Trp	Trp
		805					810				815				
Lys	Asp	Asn	Lys	Asn	Lys	Thr	Trp	Glu	Ile	Gly	Lys	Val	Ile	Thr	Trp
		820				825				830					
Gly	Arg	Gly	Phe	Ala	Cys	Val	Ser	Pro	Gly	Glu	Asn	Gln	Leu	Pro	Val
	835					840				845					
Trp	Ile	Pro	Thr	Arg	His	Leu	Lys	Phe	Tyr	Asn	Glu	Pro	Ile	Arg	Asp
	850					855				860					
Ala	Lys	Lys	Ser	Thr	Ser	Ala	Glu	Thr	Glu	Thr	Ser	Gln	Ser	Ser	Thr
	865					870			875			880			
Val	Asp	Ser	Gln	Asp	Glu	Gln	Asn	Gly	Asp	Val	Arg	Arg	Thr	Asp	Glu
	885					890				895					
Val	Ala	Ile	His	Gln	Glu	Gly	Arg	Ala	Ala	Asn	Leu	Gly	Thr	Thr	Lys
		900				905				910					
Glu	Ala	Asp	Ala	Val	Ser	Tyr	Lys	Ile	Ser	Arg	Glu	His	Lys	Gly	Asp
		915				920				925					
Thr	Asn	Pro	Arg	Glu	Tyr	Ala	Ala	Cys	Ser	Leu	Asp	Asp	Cys	Ile	Asn
	930					935				940					
Gly	Gly	Lys	Ser	Pro	Tyr	Ala	Cys	Arg	Ser	Ser	Cys	Ser	Ala		
	945					950				955					

<210> 75
<211> 12366
<212> DNA

Substitute Sequence Listing_USSN 10587032_PP019482.007

<213> Human endogenous retrovirus, K family (HERV-K), Located at 22q11.2

<400> 75

tgtggggaaa	agaaagagag	atcagactgt	tactgtgtct	atgttagaaag	aaatagacat	60
aagagactcc	attttgttct	gtactaagaa	aaattcttct	gctttgagat	gctgttaatc	120
tgtAACCTA	cccccaaccc	tgtgtcaca	gaaacagggt	ctgtgttgac	tcaaggttta	180
atggatttcg	ggctgtgcag	gatgtgctt	gttAAACAAA	tgcttgaagg	cagcaagctt	240
gttaagagtc	atcaccactc	cctaATCTCA	agtaAGCAGG	gacacaAAACA	ctgcggagg	300
cccgaggagc	ctctgcctag	gaaAGCCAGG	tgttgtccaa	ggTTTCTCCC	catgtgacag	360
tctgaaatat	ggccttgg	gaagggaaag	acctgactgt	cccctggccc	gacaccgt	420
aagggtctgt	gctgaggatt	agtaAAAGAG	gaaggaaaggc	ctcttgcag	ttgagataag	480
aggaaggcat	ctgtctcctg	ctcatccctg	ggcaatggaa	tgtcttggtg	taaAGCCTGA	540
ttgtatATGC	catctactga	gatAGGGAGAA	aactgCCtta	gggctggagg	tgggacatgc	600
tggcggcaat	actgctctt	aaggcattga	gatgtttatg	tatATGCA	tcaAAAGCAC	660
agcactttt	tcttacctt	gtttatgt	cagagacatt	tgttcacatg	ttttccTGT	720
ggccCTCTCC	ccactattac	cctattgtcc	tgccacatcc	ccctctccga	gatggtagag	780
ataatgatca	ataaatactg	agggAACTCA	gagaccgggt	cggcgggggt	cctccatATG	840
ctgagcggc	gtccccctggg	cccactttt	tttctctata	cttgtctct	gttgtcttc	900
ttttctcaag	tctctcggt	cacctgagga	gaaatgccc	cagctgtgg	ggcgcaggcc	960
actccatctg	gtgcccAACG	tggatgttt	tctctaggt	gaagggactc	tcgagtgtgg	1020
tcattgagga	caagtcaacg	agagATCCC	gagtacgtct	acagtgagcc	tttggttaag	1080
cttggggcgt	cggaagaAGC	cagggttaat	ggggcaAACT	aaaAGTAAG	tctctcattc	1140
cacctgatga	gaaACACCCA	gagggttgga	ggggcaggcc	acccttcag	ggtagggtcc	1200
cctccatgca	gaccatAGAG	cacagggtgt	ccccAAAGAG	gagcagagag	aaggaggag	1260
agggcccacg	agagACTTGG	aaatGAATGG	caggATTtta	ggcgcgtggc	ttgggttcgg	1320
ggcacCTGGC	ctttcTTTGT	gtatttctcc	tactgtctgc	ctaactattt	aatacaataa	1380
aagaaaaACCA	gccccTGGT	cttgggtgt	ttccaccctc	ccgggtcccc	gctggctgcc	1440
tggcttccTC	ccgcagctcc	tgctgtgt	gtatgtgt	gtgtgtgcac	atctgtgggg	1500
cgtatgtgt	ttcgtcttt	taattgaggc	tgcagagtgg	agagAGCAGG	gttttctct	1560
ggggacCCAG	agagaaggag	gcgtttcac	cacagccaa	caggcagga	ccccagcacc	1620
cgggacCCAG	cgggactttt	ccaaggggat	ggacctggct	ggggcAcGCG	gctgttGtg	1680
tagggAAAAG	aaAGAGAGAT	cacactgtt	ctgtgtctat	gtagAAAAGG	aagacataaa	1740
ctccatttt	agctgtacta	agaaaaattt	tttgcctt	acctgctgtt	aacctgtAAC	1800
tgtagcccc	accctgtgt	caaAGAAACA	tgtgtctat	ggaatcaagg	ttaagggt	1860
caagggtgt	acaggatgt	ccttggtaac	aatgtgtt	caggcagtat	gttgttAAAG	1920
agtcatGCC	atttccatt	ctccatttaat	caggggcacg	atgcactgcg	gaaAGCCACA	1980
gggacCTCTG	cccgagAAAG	cctgggtatt	gtccaaggct	tccccccact	gagacAGCCT	2040
gagatacggc	ctcgTGGGA	gggAAAGACC	tgaccgtccc	ccagccccac	acccgtAAAG	2100
ggTCTGTGT	gaggaggatt	agtaAAAGGG	gaaggcctct	tgcagtgtgg	ataAGAGGA	2160
ggcCTCCGTC	tctgcatgt	ccttgggaat	ggaatgtgt	gtgttAAAC	ccgatAGTAC	2220
attcCTCTA	ttcgtgagaga	agaaaaaccac	cctgtggctg	gaggatgt	atgctAGCGG	2280
caatgtgt	ctgttactct	ttgcttACACT	gagatgttt	gttggagaga	agcataAAATC	2340
tggcCTATGT	gCACATCTGG	gcacAGAAAC	tccccTTGAA	cttgcacac	agattcTTT	2400
gttcacatgt	tttcctgt	accttctccc	cactatGCC	ctgttctccc	accgcattcc	2460
ccttgcTGT	atagtAAAAA	tagtatctg	tagatACAA	gggAAACTCAG	agaccatGGC	2520
cggtgacat	cctccgtac	ctgagcgtg	gtccccTGGG	cccattgttc	tttctctata	2580
ctttgtctct	gtgtcttatt	tctttcctca	gtctctcatc	cctccTgacg	agaaataACCC	2640
acaggtgtgg	aggggCTGGC	cccCTTCATC	tgatGCCAA	tgtgggtGCC	tttctctagg	2700
gtgaaggtac	tctacagtgt	ggtcattgag	gacaAGTTGA	cgagAGAGTC	ccaaAGTACGT	2760
ccacggTCAG	ccttgcggta	agcttgtgt	cttagAGGAA	cccAGGGTA	cgatggggca	2820
aactgaaaAGT	aaatATGCCT	tttatCTCAG	tttttattAA	attcttttAA	gaagAGGGGG	2880
agttAGAGCT	tctacAGAAA	atctaattac	gctatttca	acaatAGAAC	aattCTGCCC	2940
atggTTCCa	gaacAGGGAA	ctttagatct	aaaAGATTG	gaaaaAAATTG	gcaaAGAAATT	3000
aaaacaAGCA	aatAGGGAAG	gtAAAATCAT	cccacttaca	gtatGGAATG	attgggCCAT	3060
tattaaAGCA	actttAGAAC	catttCAAAC	aggAGAAAGAT	attgtttcAG	tttctgtatc	3120
ccctAAAGC	tgtgtAAACAG	attgtGAAGA	agaggcAGGG	acagaATCCC	agcaAGGAAC	3180
ggAAAGTTCA	cattgtAAAT	atgttagcaga	gtctgtat	gctcAGTCAA	cgcaAAATGT	3240
tgactacAGT	caattacAGG	agataatATC	ccctgtat	tcaAAATTGG	ggGAAGGAGG	3300
tccagaATCA	ttggggccat	cagAGCCTAA	accacgtcg	ccatcaACTC	ctcCTCCGT	3360
ggTCAGATG	cttgcataACAT	tacaACCTCA	aacgcAGGTT	agacaAGCAC	aaACCCCAAG	3420
agAAAATCAA	gtAGAAAGGG	acagAGTCTC	tatcccggca	atgccaACTC	agatacAGTA	3480
tccacaatAT	cAGCCGTTAG	aaaATAAGAC	ccaACCGCTG	gtagtttATC	aataCCGGCT	3540
gccaaccgag	cttcagtatc	ggcctcTTTC	agaggttAA	tacagacCTC	aagcgggtgt	3600

Substitute Sequence Listing_USSN 10587032_PP019482.007

tcctgtgcca	aatagcacgg	caccatacca	gcaacccaca	gcgatggcgt	ctaattcacc	3660
agcaacacag	gacgcggcgc	tgtatcctca	gccgcccact	gtgagactta	atcctacagc	3720
atcacgtag	ggacagggtg	gtgcaactgca	tgcagtcatt	gatgaagcca	aaaaacaggg	3780
cgatcttag	gcatggcgtt	tcctgttaat	tttacaactg	gtacaggccg	gggaagagac	3840
tcaagtagga	gcmcctgccc	gagctgagac	tagatgtcaa	ccttcacca	tgaaaatgtt	3900
aaaagatata	aaggaaaggag	ttaaacaata	tggatccaa	tccccttata	taagaacatt	3960
attagattcc	attgctcatg	gaaatagact	tactccttat	gactgggaaa	ttttggccaa	4020
atcttccctt	tcatcctctc	agtatctaca	gtttaaaacc	tggtgattg	atggagtaca	4080
agaacaggta	cgaaaaaaatc	aggctactaa	gcccactgtt	aatatagacg	cagaccaatt	4140
gttaggaaca	ggtccaaatt	ggagcaccat	taaccaacaa	tcagtatgc	agaatgaggc	4200
tattgaacaa	gtaaggggcta	tttgccctag	ggcctgggaa	aaaattcagg	acccaggaac	4260
agctttccct	attaattcaa	ttagacaagg	ctctaaagag	ccatatccgt	actttgtggc	4320
aagattacaa	gatgctgctc	aaaagtctat	tacagatgac	aatgcccga	aagttattgt	4380
agaattaatg	gcctatgaaa	atgcaaatcc	agaatgtcg	tcgcccataa	agccataaa	4440
aggaaaaagtt	ccagcaggag	ttgatgtaat	tacagaatat	gtgaaggcgtt	gtgatggat	4500
tggaggagct	atgcataagg	caatgctaatt	ggctcaagca	atgagggggc	tcactctagg	4560
aggacaagtt	agaacatttgc	ggaaaaaaatg	ttataatgtt	ggtcaatcgt	gtcatctgaa	4620
aaggagttgc	ccaggcttaa	ataaacagaa	tataataat	caagcttata	cagcaaaaaa	4680
taaaaagcc	tctggctgt	gtccaaaatg	tggaaaagca	aaacattggg	ccaatcaatg	4740
tcattctaaa	tttgataaaag	atggcaacc	attgtcttga	aacaggaaga	ggggccagcc	4800
tcagggcccc	caacaaactg	gggcattttcc	agttaaactg	tttgttccctc	agggtttca	4860
aggacaacaa	cccctacaga	aaataccacc	acttcaggag	gtcagccaaat	tacaacaatc	4920
caacagctgt	cccgcccac	agcaggcagc	accgcagtag	atttatgttc	cacccaaatg	4980
gtctttttac	tccctggaaa	gcccccacaa	aagattctta	gaggggtata	tggcccctg	5040
ccagaaggga	gggttaggcct	ttgagggaga	tcaagtctaa	atttgaaggg	agtccaaatt	5100
catactgggg	taatttatttc	agattataaa	gggggaattt	agtttagtgc	cagtcact	5160
gttccccgg	gtgccaatcc	aggtataga	attgctcaat	tactgctttt	gccttatgtt	5220
aaaattgggg	aaaacaaaaa	ggaaagaaca	ggagggttt	gaagtaccaa	ccctgcagga	5280
aaagctgctt	attgggctaa	tcaggtctca	gaggatagac	ccgtgtgtac	agtcactatt	5340
cagggaaaga	gtttgaagga	tttagtggata	cccaggctga	tgttctgtc	atcggcatag	5400
gtactgcctc	agaagtgtat	caaagtccca	tgattttaca	ttgtccagga	tctgataatc	5460
aagaaagtac	ggttcagct	gtgatcactt	cattccaatc	aatttatggg	gccgagactt	5520
gttacaacaa	tggcatgcag	agattactat	cccagccctc	ctatacagcc	ccaggaataa	5580
aaaaatcatg	actaaaatgg	gatagctccc	taaaaagggaa	ctagggaaaga	agtccaaatt	5640
gaggctgaaa	aaaatcaaaa	aagaaaagga	atagggcatc	cttttttagga	gcggtcactg	5700
tagagccctc	aaaacccatt	ccatttaactt	ggggggaaaaa	aaaacaactg	tatgttaaat	5760
cagcagcgct	tccaaacaa	aaactggagg	ctttagcattt	attagcaag	aaacaattag	5820
aaaaaggaca	ttgagccctc	attttcgctt	tggatttctg	ttttaatttc	agaaaaaaatc	5880
cggcagatgg	cgtataatgc	cgttaatcaa	ccccatgggg	cttccttccacc	ccgggttgc	5940
tctccagcca	tgttccccctt	taatttataat	tgatctgaag	gattgctttt	ttaccattcc	6000
tctggaaaaa	caggattttt	aaaaatttgc	ttttaccaca	ccagcctaaa	taataaagaa	6060
ccagccacca	ggtttcagtg	gaaagtattt	cctcaggggaa	tgcttaatag	ttcaactatt	6120
tgtcagctca	agctctgcaa	ccagtttagag	acaagttttc	agactgttac	atcggttact	6180
atgttgcata	tttgtgtgt	gcagaaaacga	gagacaattt	aattgaccgt	tacacatttc	6240
tgcagacaga	ggttgccaac	gcgggactgt	caataacatc	tgataagatt	caaacctcta	6300
ctccttccg	ttacttggga	atgcaggtag	aggaaaggaa	aattaaacca	aaaaaaatag	6360
aaataagaaa	agacacattt	aaagcattaa	atgagttca	aaagttgcta	ggagatacta	6420
attggatttgc	gagatattaa	ttggattttgg	ccaaactctag	gcatttcctac	ttatgcccatt	6480
tcaaattttgt	tcttttctt	aagaggggac	tcggaattaa	atagtgaaag	aacgttaact	6540
ccagaggcaa	ctaaagaaat	taaattaatt	gaagaaaaaa	ttcggtcagc	acaagtaat	6600
agaatagatc	acttggcccc	actccaaattt	ttgatttttgc	ctactgcaca	ttccctaaaca	6660
ggcatcatttgc	ttcaaaatac	agatcttgc	gagtggcctt	tccttcctca	cagtacaatt	6720
aagacttttgc	cattgtactt	ggatccaaatg	gctacattaa	ttggtcaggg	aagattatga	6780
ataataacat	tgtgtggaaa	tgaccaggat	aaaatcactg	ttccttccaa	caagcaacag	6840
gttagacaag	cctttatcaa	ttctggtgca	tggcagatttgc	gttttgcgcg	ttttgtggaa	6900
attatttgcata	atcggttacc	caaaaacaaaa	atcttccagt	ttttaaaattt	gactacttgg	6960
attttaccta	aagtttacca	acataaggcct	ttaaaaaatg	ctctggcagt	gtttactgtat	7020
ggttccagca	atggaaaaatg	ggcttacacc	gggcctaaaa	aataggtcat	aaaaactcag	7080
tatcacttgc	ctcaaagagc	agagttggtt	gccgttacca	cagtgttac	aagattttaa	7140
tcagtcatttgc	aacattgtat	cagattctgc	atatgttagt	caggttacaa	aggatattga	7200
gagagcccta	atcaaataca	ttatggatga	tcagtttaac	ccgcttgcata	atttgcata	7260
acaaaatgtt	agaaaaaaagaa	atttcccattt	ttatattact	catattcgag	cacacactaa	7320
tttaccagggtt	cctttaacta	aagcaaata	acaagctgac	ttgttagtat	catctgcatt	7380

Substitute Sequence Listing_USSN 10587032_PP019482.007

catggaagca	caagaacttc	atgccttgac	tcatgttaat	gcaataggat	taaaaaataa	7440
atttgatatac	acatggaaac	agacaaaaaa	tattgtacaa	cattgcaccc	agtgtcagat	7500
tctacacctg	gccactcagg	aggcaagagt	taatcccaga	ggcttatgtc	ctaattgtgtt	7560
atggcaaatg	gatgtcatgc	acgtaccttc	atttggaaaa	ttgtcatttg	tccatgtgac	7620
agttgatact	tattcacatt	tcatatggc	aacctgcagg	acaggagaaa	gtactttcca	7680
tgttaaaaaga	catttattat	cttgggttcc	tgtcatggg	gttccagaaa	aagttaaaac	7740
agacaatggg	ccaggttact	gtagtaaagc	agttcaaaaa	ttcttaaattc	agtggaaaat	7800
tacacataca	ataggaattc	tctataattc	ccaaggacag	gccataattg	aaagaactaa	7860
tagaacactc	aaagctcaat	tggtaaaaca	aaaaaaagga	aaagacagga	gtataacact	7920
ccccagatgc	aacttaatct	agcactctat	actttaaatg	ttttaaacat	ttatagaat	7980
cagaccacta	cctctgcaga	acaacatctt	actggtaaaa	ggaacagccc	acatgaagga	8040
aaactgattt	ggtggaaaga	taataaaaat	aaaacatggg	aatggggaa	ggtgataacg	8100
tgggggagag	gttttgcttg	tgttgcacca	ggagaaaatc	agcttccctgt	ttggataaccc	8160
actagacatt	taaagttcta	caatgaactc	actggagatg	caaagaaaag	tgtggagatg	8220
gagacacccc	aatcgactcg	ccaggtaaac	aaaatggtga	tatcagaaga	acagaaaaag	8280
ttgccttcca	tcaaggaagc	agagttgcca	atataggcac	aattaaagaa	gctgacacag	8340
tttagctaaaa	aaaaaaagct	agagaataca	aaggtgacac	caactccaga	gaatatgctg	8400
cttgcagctc	tgtatgattgt	atcaacggtg	gtaagtcttc	ccaagtctgc	aggagcagct	8460
gcagctaatt	atacttactg	ggcctatgtg	ccttcccac	ccttaattcg	ggcagttaca	8520
tagatggata	atccttattga	agtagatgtt	aataatagt	catgggtgcc	tggcccccaca	8580
gatgactgtt	gcccgtccca	acctgaagaa	ggaatgtga	tgaattttc	cattgggtat	8640
ctttatccct	ctgtttgcct	agggaggca	ccaggatgt	taatgcctac	aacccaaaat	8700
tggtttgtag	aagtacctac	agtcaatgt	accagtagat	ttacttatac	catggtaagt	8760
ggaatgtcac	agataaaataa	tttacaggac	ccttcttatac	aaagatcatt	acaatgttagg	8820
cctaagggga	aggctgccc	caaggaaatt	cccaaagaat	caaaaagccc	agaagtctta	8880
gtctcgagg	aatgtgtggc	tgatactgca	gtgttagtaca	aaacaatgaa	ttttgaacta	8940
tgatagactg	ggtcccttga	ggccaattat	atcataactg	tacaggccag	actcattcat	9000
gttcacaggc	cccatccatc	tggcccatta	atccagccta	tgacgggtat	gtaactgaaa	9060
ggctggacca	gttttataga	aggttagaat	cactctgtcc	aaggaaatgg	ggtggaaaagg	9120
gaatttcatc	accttgacca	aagtttagtcc	tgttactgtt	cctgaacatc	cagaatttagg	9180
aagcttactg	tggcctcaca	ccacattaga	atttggttcg	gaaatcaagc	tataggaaca	9240
agagatcgta	agtcatattt	tactatcaac	ctaaatttca	gtctgacaat	tcctttgca	9300
aattgtgtaa	aactccctt	tattgtctgt	tgttagaaaa	acatagttat	taaacctgat	9360
tcccaaacc	taatctgtga	aaattgtgga	atgtttactt	gcattgattt	gacttttaat	9420
tggcagcacc	gtattctact	aggaagagca	agagagggtg	tgtggatct	tgtgtccatg	9480
gaccgaccat	ggggaggcttc	gctatccatc	catattttaa	cggaagtatt	aaaaggaatt	9540
ctaactagat	ccaaaagatt	cattttact	ttgatggcag	tgatttatggg	cctctattgc	9600
gtcacagcta	ctgctgcggc	tgctggattt	gctttacact	cctctgttca	aactgcagaa	9660
tacgttaaatg	attggcaaaa	gaattcctca	aaattgtgga	atttctcgat	ccaaatagat	9720
caaaaattgg	caaaccaaaat	taatgtat	agacaaactg	tcattttggat	gggagaggct	9780
catgagctt	gaatatcttt	ttcagttacg	atgtgactgg	aatacatcg	atttttgtgt	9840
tacaccacaa	gcctataatg	agtctgagca	tcactggac	atggttagat	gccatctgca	9900
aggaggagaa	gataatctt	ctttagacat	ttcaaaaattt	aaagaattttt	ttttttcttt	9960
gagacagagt	ctcgctctgt	cgcccaggct	ggagtgcagt	ggcgtgatct	cagctcactg	10020
caagttccgc	ctcctgggtt	tacaccattc	tcctgcctca	gcctcccaag	tagttggac	10080
tacaggagcc	caccaccatg	cctggctaat	tttttttggg	tttttaatag	agatggagtt	10140
tcaccgttt	agccaggatg	gtctcgatct	cctgaccctt	tgtctgccc	accttggcct	10200
cccaaagtgc	tgggattaca	gtcgtgagcc	accgtgccc	gccaagaaaa	aatttttgag	10260
gcataaaaag	cccatataaa	tttggtgcca	ggaacggaga	caatcgtaa	agctgctgat	10320
agcctcacaa	atcttaagcc	agtcaatttt	gttaaaagca	tcagaagttt	cactattgt	10380
aatttcatata	taatccctgt	atgcctgttc	tgtctgttgt	tagtctacag	gtgtatccag	10440
cagctccaaa	gagacagcaa	ccagcaagaa	tgggccatag	tgacgatgtt	gtttttgtca	10500
aaaagaaaaag	ggggggatata	gtaagggaaa	gagagatcg	actttcaatg	tgtctatgt	10560
gaaaaggaag	acataagaaa	ctccatattt	atctgtacta	agaaaaattt	ttttgccttg	10620
agatgtgtt	aatctgtac	tttagcccc	accctgtgt	cacggaaaca	tgtgtctgt	10680
ggttaagg	atctagggtt	gtgcaggatg	taccttggta	acaatatgtt	tgcaggcagt	10740
atgttttgta	aaagtcatcg	ccatttctca	ttctcgat	accaggggct	caatgcactg	10800
tggaaagcc	caggAACCTC	tgcccaagaa	agcctggctg	ttgtggaaag	tcaggaccc	10860
cgaatggagg	gaccagctgg	tgctgcacca	ggaaacataa	atttgtgaaga	tttcttgac	10920
atttatcagt	tttccaaaattt	aatactttt	taattttca	caccgttctt	actttatct	10980
cttaatccctg	ttatctttgt	aagctgagga	tatacgctac	ctcaggacca	ctattgtaca	11040
aattgattgt	aaaacatgtt	cacatgttt	tgaacaatat	gaaatcgatg	caccttggaaa	11100
atgaacagaa	taacagtgtat	tttagggaaac	aaaggaagac	aaccataagg	tctgactgcc	11160

Substitute Sequence Listing_USSN 10587032_PP019482.007

tgaggggctcg	ggcaaaaagc	catattttc	ttcttgcaga	gaggctataa	atggacgtgc	11220
aagttaggaga	gatattgcta	aatttttc	ctagcaagga	atataatact	aagaccctag	11280
ggaaagaatt	gcatttcctgg	ggggagggtct	ataaacgccc	gctctggag	tgtctgtcct	11340
atgtggttga	gataaggact	gagatacgcc	ctggctccct	gcagtaccct	caggcttact	11400
aggattggga	aaccccgatc	ctggtaaatt	tgaggtcagg	ccggttcttt	gctctgaacc	11460
ctgtttctg	ttaagatgtt	tatcaagaca	atacatgcac	cgctgaacat	agacccttat	11520
caggagtttc	tgatttgtct	ctggccctgt	ttcttcagaa	gcatgtcatac	tttgctctgc	11580
cttctgccct	ttgaagcatg	tgatctttgt	gacctactcc	ctgttcatac	acccctcccc	11640
ttttaaaatc	cctaataaaaa	acttgctgg	tttggggctc	aggggggcat	catggaccta	11700
ccaatacgtg	atgtcacccc	cggtggccca	gctgtaaaat	tccttcttt	atactcttat	11760
ttctcagacc	agctgacact	tagggaaaat	agaaaagaacc	tatgtgaaa	tattggaggc	11820
gggttcccccc	gataacctggg	tatttccaa	ggtttccctt	gctgaggagg	attagtaaaa	11880
ggaatgcctc	catctcctgc	atgtccctgg	gaacagaatg	ttcccaccaa	ccaccctgtg	11940
gctggaggcg	ggatatgctg	gcagcaatgc	tgctctatta	ctcttgcata	cactgagatg	12000
tttgggtgga	gagaaggata	aatctggcct	atgtgcacat	ctgggcacag	cacccttcctt	12060
tgaacttatt	tgtgacacag	attcctttgc	tcacgtttc	ctgttgaactt	tctcaccact	12120
cacccttattc	tccgtgtggc	ttcgccctgc	ggagatagtg	aaaatagtaa	taaatactga	12180
gggaacttcag	actgagggaa	ctcagactgg	gcagaccggg	gccagtgtgg	gtcctccata	12240
tgctgagcgc	cgggtccctg	ggcccaactgt	tctttctcta	tactttgtct	ctgtgcctta	12300
ttttctcagt	ctctcattcc	acctgtatgag	aaatacccac	aggtgtggag	gggctggccc	12360
ccttca						12366

<210> 76

<211> 2148

<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 76

atggggcaaa	ctgaaagttaa	atatgcctct	tatctcagct	ttattaaaat	tcttttaaga	60
agagggggag	tttagagcttc	tacagaaaat	ctaattacgc	tatttcaaac	aatagaacaa	120
ttctgcccatt	ggtttccaga	acagggaaact	ttagatctaa	aagattggga	aaaaattggc	180
aaagaattaa	aacaagcaaa	tagggaaagg	aaaatcatcc	cacttacagt	atggaaatgat	240
tgggccattt	ttaaagcaac	tttagaacca	tttcaaacag	gagaagatat	tgtttcagtt	300
tctgatgccc	ctaaaagctg	tgtaacagat	tgtgaagaag	aggcagggac	agaatcccag	360
caaggaacgg	aaagttcaca	ttgtaaaat	gttagcagat	ctgtaatggc	tcagtcaacg	420
caaaatgttg	actacagtca	attacaggag	ataatatacc	ctgaatcatc	aaaattgggg	480
gaaggaggtc	cagaatcatt	ggggccatca	gagcctaaac	cacgatcgcc	atcaactcct	540
cctccctgtt	ttcagatgcc	tgtaacacca	caacctcaaa	cgcaggttag	acaagcacaa	600
accccaagag	aaaatcaagt	agaaaaggac	agagtctta	tcccgcaat	gccaactcag	660
atacagtatc	cacaatatca	gccgttagaa	aataagaccc	aaccgtctgt	agtttatcaa	720
taccggctgc	caaccggatc	tcagtatcg	cctccttcag	aggttcaaa	cagacctcaa	780
gcggtgtgtc	ctgtgccaat	tagcacggc	ccataccgc	aaccacacgc	gatggcgtct	840
aattcaccag	caacacagg	cgcggcgtg	tatcctcagc	cgcccactgt	gagacttaat	900
cctacagcat	cacgtatgg	acagggttgt	gcactgcatt	cagtcattga	tgaagccaga	960
aaacagggcg	atcttggggc	atggcggttc	ctggtaat	tacaactgtt	acaggccggg	1020
gaagagactc	aagttaggagc	gcctgcccga	gctgagacta	gatgtgaacc	tttcaccatg	1080
aaaatgttaa	aagatataaa	ggaaggagtt	aaacaatatg	gatccaactc	cccttatata	1140
agaacattat	tagattccat	tgctcatgga	aatagactt	ctccttatga	ctggaaatt	1200
ttggccaaat	cttccctttc	atcctctcag	tatctacat	ttaaaacctg	gtggattgtat	1260
ggagtacaag	aacagggtac	aaaaaatcag	gctactaagc	ccactgtttaa	tatagacgca	1320
gaccaattgt	taggaacagg	tccaaattgg	agcaccatta	accaacaatc	agtgtatgcag	1380
aatgaggct	ttgaacaagt	aagggtatt	tgcctcaggg	cctggggaaa	aattcaggac	1440
ccaggaacag	ctttccctat	taattcaatt	agacaaggct	ctaaagagcc	atatcctgac	1500
tttggggcaa	gattacaaga	tgctgctaa	aagtcttata	cagatgacaa	tgcccggaaa	1560
gttattgttag	aattaatggc	ctatgaaaat	gcaaattccag	aatgtcagtc	ggccataaaag	1620
ccattaaaag	gaaaagttcc	agcaggagtt	gatgtat	cagaatatgt	gaaggcttgc	1680
gatgggattt	gaggagctat	gcataaggca	atgctaattgg	ctcaagcaat	gagggggctc	1740
actctaggag	gacaagtttg	aacatttggg	aaaaaatgtt	ataattgtgg	tcaaattcggt	1800
catctgaaaa	ggagttgccc	agtcttaat	aaacagaata	taataatca	agcttattaca	1860
gaaaaaaaaata	aaaagccat	tggccgtgt	ccaaaatgt	gaaaaggaaa	acattggccc	1920
aatcaatgtc	attctaaatt	tgataaggat	ggcaaccat	tgtcgggaaa	caggaagagg	1980
ggccagcctc	aggccccccca	acaaaatggg	gcattccag	ttcaactgtt	tgttcctcag	2040
gttttcaag	gacaacaacc	cctacagaaa	ataccaccac	ttcaggaggt	cagccaatta	2100

Substitute Sequence Listing_USSN 10587032_PP019482.007

caacaatcca acagctgtcc cgcgccacag caggcagcac cgcatgtaa

2148

<210> 77
<211> 2151
<212> DNA
<213> Artificial Sequence

<220>
<223> Manipulated Gag

<400> 77

atggggccaga	ccgagagcaa	gtacgccagc	tacctgagct	tcatcaagat	cctgctgcgc	60
cgcggcggcg	tgcgcgcccag	caccgagaac	ctgatcaccc	tgttccagac	catcgagcag	120
ttctgccccct	ggttccccga	gcagggcacc	ctggacactg	aggactggga	gaagatcggc	180
aaggagctga	agcaggccaa	ccgcgagggc	aagatcatcc	ccctgaccgt	gtggAACgcac	240
tgggccatca	tcaaggccac	cctggagccc	ttccagaccc	gcgaggacat	cgtgagcgtg	300
agcgcacgccc	ccaagagctg	cgtgaccgac	tgcgaggagg	aggccggcac	cgagagccag	360
cagggcaccg	agagcagcca	ctgcaagtac	gtggccgaga	gcgtgatggc	ccagagcacc	420
cagaacgtgg	actacagcca	gctgcaggag	atcatctacc	ccgagagcag	caagctggc	480
gagggcggcc	ccgagagcct	gggccccagc	gagcccaagc	cccgcagccc	cagcaccccc	540
ccccccgtgg	tgcagatgcc	cgtgaccctg	cagccccaga	cccaagggtgcg	ccaggcccag	600
accccccgcg	agaaccaggt	ggagcgcgac	cgcgtgagca	tccccgccat	gcccaccagg	660
atccagtacc	cccagtacca	ggccgtggag	aacaagaccc	agccccctgg	ggtgtaccag	720
taccgcctgc	ccacccgagct	gcagtaccgc	ccccccagcg	aggtgcagta	ccgccccccag	780
gcccgtgtcc	ccgtgccccaa	cagcaccgccc	cccttaccagc	agcccaccgc	catggccagc	840
aacagccccg	ccacccagga	cgccgcctcg	taccccccagc	ccccccaccgt	gcccctgaac	900
cccaccgcga	gccgcagcgg	ccagggcggc	gccctgcacg	ccgtgatcga	cgaggcccgc	960
aagcagggcgc	acctggaggc	ctggcgcttc	ctgggtatcc	tgcagctgg	gcaggccggc	1020
gaggagaccc	aggtggggcgc	ccccggccgc	gccgagaccc	cttgcgagcc	cttcaccatg	1080
aagatgtga	aggacatcaa	ggagggcgtg	aagcgtacg	gcagcaacag	cccctacatc	1140
cgcaccctgc	tggacagcat	cgcacccacggc	aaccgcctga	ccccctacga	ctgggagatc	1200
ctggccaaga	gcagccttag	cagcagcccg	tacctgcgt	tcaagacctg	gtggatgcac	1260
ggcgtgcagg	agcagggtgc	caagaaccag	gccaccaagc	ccaccgtaa	catcgacgcc	1320
gaccagtgc	tgggcaccgg	ccccaaactgg	agcaccatca	accaggag	cgtgatgcag	1380
aacgaggcca	tcgagcagggt	gcgcgcccattc	tgcctgcgcg	cctggggcaa	gatccaggac	1440
ccggcaccg	ccttccccat	caacagcatc	cgccaggggca	gcaaggagcc	ctaccccgac	1500
ttcgtggccc	gcctgcagga	cgccggccag	aagagcatca	ccgacgacaa	cgcccgcag	1560
gtgatgtgg	agtgtatggc	ctacgagaac	gccaaccccg	agtgccagag	cgccatcaag	1620
cccctgaagg	gcaagggtgcc	cgccggcgtg	gacgtatca	ccgagtacgt	gaaggcctgc	1680
gacggcatcg	gcggcgccat	gcacaaggcc	atgctgtatgg	cccaggccat	gcccggcctg	1740
accctgggcg	gccagggtcg	caccttgcgc	aagaagtgt	acaactgcgg	ccagatcgcc	1800
cacctgaagc	gcagctgccc	cgtgctgaac	aagcagaaca	tcatcaacca	ggccatcacc	1860
gccaagaaca	agaagcccg	cgccctgtgc	cccaagtgcg	gcaagggcaa	gcactgggcc	1920
aaccagtgtcc	acagcaagtt	cgacaaggac	ggccagcccc	tgagcggcaa	ccgcaagcgc	1980
ggccagcccc	aggcccccca	gcagaccggc	gccttccccg	tgcaagctgtt	cgtgccccag	2040
ggcttccagg	gccagcagcc	cctgcaagaag	atccccccccc	tgcaaggcgt	gagccagctg	2100
cagcagagca	acagctgccc	cgccccccag	caggccgccc	cccaqqctta	a	2151

<210> 78

<211> 715

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 78

Met Gly Gln Thr Glu Ser Lys Tyr Ala Ser Tyr Leu Ser Phe Ile Lys
1 5 10 15

Ile Leu Leu Arg Arg Gly Gly Val Arg Ala Ser Thr Glu Asn Leu Ile
20 25 30

Thr Leu Phe Gln Thr Ile Glu Gln Phe Cys Pro Trp Phe Pro Glu Gln
35 40 . 45

Substitute Sequence Listing_USSN 10587032_PP019482.007

Gly Thr Leu Asp Leu Lys Asp Trp Glu Lys Ile Gly Lys Glu Leu Lys
50 55 60

Gln Ala Asn Arg Glu Gly Lys Ile Ile Pro Leu Thr Val Trp Asn Asp
65 70 75 80

Trp Ala Ile Ile Lys Ala Thr Leu Glu Pro Phe Gln Thr Gly Glu Asp
85 90 95

Ile Val Ser Val Ser Asp Ala Pro Lys Ser Cys Val Thr Asp Cys Glu
100 105 110

Glu Glu Ala Gly Thr Glu Ser Gln Gln Gly Thr Glu Ser Ser His Cys
115 120 125

Lys Tyr Val Ala Glu Ser Val Met Ala Gln Ser Thr Gln Asn Val Asp
130 135 140

Tyr Ser Gln Leu Gln Glu Ile Ile Tyr Pro Glu Ser Ser Lys Leu Gly
145 150 155 160

Glu Gly Pro Glu Ser Leu Gly Pro Ser Glu Pro Lys Pro Arg Ser
165 170 175

Pro Ser Thr Pro Pro Pro Val Val Gln Met Pro Val Thr Leu Gln Pro
180 185 190

Gln Thr Gln Val Arg Gln Ala Gln Thr Pro Arg Glu Asn Gln Val Glu
195 200 205

Arg Asp Arg Val Ser Ile Pro Ala Met Pro Thr Gln Ile Gln Tyr Pro
210 215 220

Gln Tyr Gln Pro Val Glu Asn Lys Thr Gln Pro Leu Val Val Tyr Gln
225 230 235 240

Tyr Arg Leu Pro Thr Glu Leu Gln Tyr Arg Pro Pro Ser Glu Val Gln
245 250 255

Tyr Arg Pro Gln Ala Val Cys Pro Val Pro Asn Ser Thr Ala Pro Tyr
260 265 270

Gln Gln Pro Thr Ala Met Ala Ser Asn Ser Pro Ala Thr Gln Asp Ala
275 280 285

Ala Leu Tyr Pro Gln Pro Pro Thr Val Arg Leu Asn Pro Thr Ala Ser
290 295 300

Arg Ser Gly Gln Gly Gly Ala Leu His Ala Val Ile Asp Glu Ala Arg
305 310 315 320

Lys Gln Gly Asp Leu Glu Ala Trp Arg Phe Leu Val Ile Leu Gln Leu
325 330 335

Val Gln Ala Gly Glu Glu Thr Gln Val Gly Ala Pro Ala Arg Ala Glu
340 345 350

Thr Arg Cys Glu Pro Phe Thr Met Lys Met Leu Lys Asp Ile Lys Glu
355 360 365

Gly Val Lys Gln Tyr Gly Ser Asn Ser Pro Tyr Ile Arg Thr Leu Leu
370 375 380

Substitute Sequence Listing_USSN 10587032_PP019482.007

Asp Ser Ile Ala His Gly Asn Arg Leu Thr Pro Tyr Asp Trp Glu Ile
385 390 395 400

Leu Ala Lys Ser Ser Leu Ser Ser Ser Gln Tyr Leu Gln Phe Lys Thr
405 410 415

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Lys Asn Gln Ala Thr
420 425 430

Lys Pro Thr Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Thr Gly Pro
435 440 445

Asn Trp Ser Thr Ile Asn Gln Gln Ser Val Met Gln Asn Glu Ala Ile
450 455 460

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Gly Lys Ile Gln Asp
465 470 475 480

Pro Gly Thr Ala Phe Pro Ile Asn Ser Ile Arg Gln Gly Ser Lys Glu
485 490 495

Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Ala Ala Gln Lys Ser
500 505 510

Ile Thr Asp Asp Asn Ala Arg Lys Val Ile Val Glu Leu Met Ala Tyr
515 520 525

Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys Gly
530 535 540

Lys Val Pro Ala Gly Val Asp Val Ile Thr Glu Tyr Val Lys Ala Cys
545 550 555 560

Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln Ala
565 570 575

Met Arg Gly Leu Thr Leu Gly Gly Gln Val Arg Thr Phe Gly Lys Lys
580 585 590

Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Arg Ser Cys Pro Val
595 600 605

Leu Asn Lys Gln Asn Ile Ile Asn Gln Ala Ile Thr Ala Lys Asn Lys
610 615 620

Lys Pro Ser Gly Leu Cys Pro Lys Cys Gly Lys Gly Lys His Trp Ala
625 630 635 640

Asn Gln Cys His Ser Lys Phe Asp Lys Asp Gly Gln Pro Leu Ser Gly
645 650 655

Asn Arg Lys Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
660 665 670

Pro Val Gln Leu Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Leu
675 680 685

Gln Lys Ile Pro Pro Leu Gln Gly Val Ser Gln Leu Gln Gln Ser Asn
690 695 700

Ser Cys Pro Ala Pro Gln Gln Ala Ala Pro Gln
705 710 715

Substitute Sequence Listing_USSN 10587032_PP019482.007

<210> 79
<211> 716
<212> PRT
<213> Artificial Sequence

<220>
<223> Manipulated Gag

<400> 79
Met Gly Gln Thr Glu Ser Lys Tyr Ala Ser Tyr Leu Ser Phe Ile Lys
1 5 10 15
Ile Leu Leu Arg Arg Gly Gly Val Arg Ala Ser Thr Glu Asn Leu Ile .
20 25 30
Thr Leu Phe Gln Thr Ile Glu Gln Phe Cys Pro Trp Phe Pro Glu Gln
35 40 45
Gly Thr Leu Asp Leu Lys Asp Trp Glu Lys Ile Gly Lys Glu Leu Lys
50 55 60
Gln Ala Asn Arg Glu Gly Lys Ile Ile Pro Leu Thr Val Trp Asn Asp
65 70 75 80
Trp Ala Ile Ile Lys Ala Thr Leu Glu Pro Phe Gln Thr Gly Glu Asp
85 90 95
Ile Val Ser Val Ser Asp Ala Pro Lys Ser Cys Val Thr Asp Cys Glu
100 105 110
Glu Glu Ala Gly Thr Glu Ser Gln Gln Gly Thr Glu Ser Ser His Cys
115 120 125
Lys Tyr Val Ala Glu Ser Val Met Ala Gln Ser Thr Gln Asn Val Asp
130 135 140
Tyr Ser Gln Leu Gln Glu Ile Ile Tyr Pro Glu Ser Ser Lys Leu Gly
145 150 155 160
Glu Gly Gly Pro Glu Ser Leu Gly Pro Ser Glu Pro Lys Pro Arg Ser
165 170 175
Pro Ser Thr Pro Pro Val Val Gln Met Pro Val Thr Leu Gln Pro
180 185 190
Gln Thr Gln Val Arg Gln Ala Gln Thr Pro Arg Glu Asn Gln Val Glu
195 200 205
Arg Asp Arg Val Ser Ile Pro Ala Met Pro Thr Gln Ile Gln Tyr Pro
210 215 220
Gln Tyr Gln Pro Val Glu Asn Lys Thr Gln Pro Leu Val Val Tyr Gln
225 230 235 240
Tyr Arg Leu Pro Thr Glu Leu Gln Tyr Arg Pro Pro Ser Glu Val Gln
245 250 255
Tyr Arg Pro Gln Ala Val Cys Pro Val Pro Asn Ser Thr Ala Pro Tyr
260 265 270
Gln Gln Pro Thr Ala Met Ala Ser Asn Ser Pro Ala Thr Gln Asp Ala
275 280 285

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ala Leu Tyr Pro Gln Pro Pro Thr Val Arg Leu Asn Pro Thr Ala Ser
290 295 300

Arg Ser Gly Gln Gly Gly Ala Leu His Ala Val Ile Asp Glu Ala Arg
305 310 315 320

Lys Gln Gly Asp Leu Glu Ala Trp Arg Phe Leu Val Ile Leu Gln Leu
325 330 335

Val Gln Ala Gly Glu Glu Thr Gln Val Gly Ala Pro Ala Arg Ala Glu
340 345 350

Thr Arg Cys Glu Pro Phe Thr Met Lys Met Leu Lys Asp Ile Lys Glu
355 360 365

Gly Val Lys Gln Tyr Gly Ser Asn Ser Pro Tyr Ile Arg Thr Leu Leu
370 375 380

Asp Ser Ile Ala His Gly Asn Arg Leu Thr Pro Tyr Asp Trp Glu Ile
385 390 395 400

Leu Ala Lys Ser Ser Leu Ser Ser Gln Tyr Leu Gln Phe Lys Thr
405 410 415

Trp Trp Ile Asp Gly Val Gln Glu Gln Val Arg Lys Asn Gln Ala Thr
420 425 430

Lys Pro Thr Val Asn Ile Asp Ala Asp Gln Leu Leu Gly Thr Gly Pro
435 440 445

Asn Trp Ser Thr Ile Asn Gln Gln Ser Val Met Gln Asn Glu Ala Ile
450 455 460

Glu Gln Val Arg Ala Ile Cys Leu Arg Ala Trp Gly Lys Ile Gln Asp
465 470 475 480

Pro Gly Thr Ala Phe Pro Ile Asn Ser Ile Arg Gln Gly Ser Lys Glu
485 490 495

Pro Tyr Pro Asp Phe Val Ala Arg Leu Gln Asp Ala Ala Gln Lys Ser
500 505 510

Ile Thr Asp Asp Asn Ala Arg Lys Val Ile Val Glu Leu Met Ala Tyr
515 520 525

Glu Asn Ala Asn Pro Glu Cys Gln Ser Ala Ile Lys Pro Leu Lys Gly
530 535 540

Lys Val Pro Ala Gly Val Asp Val Ile Thr Glu Tyr Val Lys Ala Cys
545 550 555 560

Asp Gly Ile Gly Gly Ala Met His Lys Ala Met Leu Met Ala Gln Ala
565 570 575

Met Arg Gly Leu Thr Leu Gly Gly Gln Val Arg Thr Phe Gly Lys Lys
580 585 590

Cys Tyr Asn Cys Gly Gln Ile Gly His Leu Lys Arg Ser Cys Pro Val
595 600 605

Leu Asn Lys Gln Asn Ile Ile Asn Gln Ala Ile Thr Ala Lys Asn Lys
610 615 620

Substitute Sequence Listing_USSN 10587032_PP019482.007
Lys Pro Ser Gly Leu Cys Pro Lys Cys Gly Lys Gly Lys His Trp Ala
625 630 635 640

Asn Gln Cys His Ser Lys Phe Asp Lys Asp Gly Gln Pro Leu Ser Gly
645 650 655

Asn Arg Lys Arg Gly Gln Pro Gln Ala Pro Gln Gln Thr Gly Ala Phe
660 665 670

Pro Val Gln Leu Phe Val Pro Gln Gly Phe Gln Gly Gln Gln Pro Leu
675 680 685

Gln Lys Ile Pro Pro Leu Gln Gly Val Ser Gln Leu Gln Gln Ser Asn
690 695 700

Ser Cys Pro Ala Pro Gln Gln Ala Ala Pro Gln Ala
705 710 715

<210> 80

<211> 6486

<212> DNA

<213> Artificial Sequence

<220>

<223> pCMVKm2.gagopt PCAV vector

<400> 80

gccgcggaat ttcgactcta ggccattgca tacgttgat ctagatcata atatgtacat 60
ttatattggc tcatgtccaa tatgaccgccc atgttgcacat tgattattga ctatgttatta 120
atagtaatca attacgggtt cattagttca tagccccat atggagttcc gcgttacata 180
acttacggta aatggcccgcc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat 240
aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc aatgggttgg 300
gtatttacgg taaaactgccc acttggcagt acatcaagt tatcatatgc caagtccgccc 360
ccctatttgc gtcataatgacg gtaaatggcc cgccctggcat tatgcccagt acatgaccc 420
acgggacttt cctacttggc agtacatcta cgtttagtgc atcgcttata ccatgggtat 480
gcgggtttgg cagtagccca atggggctgg atagcgggtt gactcacggg gatttccaag 540
tctccacccc attgacgtca atggggatgtt gttttggcac caaatcaac gggactttcc 600
aaaatgtcata aataaccccg ccccggttgc gcaaatgggc ggtaggcgtg tacgggtgg 660
ggtctatata agcagagctc gtttagtggaa ccgtcagatgc gcctggagac gccatccac 720
ctgttttgc ctccatagaa gacaccgggg ccgtatccgc ctccgcggcc gggAACGGGT 780
cattggaaacg cggattccccc gtgccaagag tgacgttaatg accgcctata gactctatag 840
gcacacccct ttggctctta tgcatgtatc actgtttttg gcttggggcc tatacaccc 900
cgcttccctt tgctataggt gatggtatacg ctttagccat aggtgtgggtt tattgaccat 960
tattgaccac tccccttattt gtgacgatac tttccattac taatccataa catggctctt 1020
tgccacaact atctcttattt gctatatgcc aataactctgt ccttcagaga ctgacacgg 1080
ctctgtatTT ttacaggatg gggcccattt tattattttac aaatttcacat atacaacaac 1140
gccgtcccccc gtgcccgcag tttttattaa acatagcgtg ggatctccac gcgaatctcg 1200
ggtagctgtt ccggacatgg gctcttctcc ggtagcggcg gagcttccac atccgagccc 1260
tggtcccattt cttccagcggtt ctcatggtcg ctcggcagct ctttgtctt aacagtggag 1320
gccagactta ggcacagcac aatgcccacc accaccatgt tgccgcacaa ggccgtggcg 1380
gttagggatgt tgcgtaaaaa tgagctcgga gattgggctc gcaccgctga cgcagatgga 1440
agacttaagg cagcggcaga agaagatgca ggcagctgag ttgttgtatt ctgataagag 1500
tcagaggtaa ctcccgttgc ggtgctgtt acgggtggagg gcagtgttagt ctgagcagta 1560
ctcggttctg cccgcgcgc caccagacat aatagctgac agactaacag actgtttctt 1620
tccatgggtc ttttctgcag tcaccgtcgt cgacgcaccat atgggccaga ccgagagcaa 1680
gtacggccagc tacctgagct tcataaagat cctgctgcgc cgcggcggcg tgcgcgccag 1740
caccgagaac ctgatcaccc tggccatgc acatcgacatgc cttctggccctt gtttccccga 1800
gcaggggcacc ctggacatgc aggactggga gaagatggc aaggagctga agcaggccaa 1860
ccgcgaggggc aagatcatcc ccctgaccgt gtggaaacgac tggccatca tcaaggccac 1920
cctggagcccc ttccagaccc gcgaggacat cgtgagcgtg agcagacccc ccaagagctg 1980
cgtgaccgac tgcgaggagg aggccggcac cgagagccag caggccaccg agagcagcca 2040
ctgcaagtac gtggccgaga gctgtatggc ccagagcacc cagaacgtgg actacagcca 2100
gctgcaggag atcatctacc ccgagagcagc caagctggc gagggcggcc ccgagagcct 2160

Substitute Sequence Listing_USSN 10587032_PP019482.007

gggccccagc	gagcccaagc	cccgacgccc	cagcaccccc	ccccccgtgg	tgcagatgcc	2220
cgtgaccctg	cagccccaga	cccagggtcg	ccaggcccag	acccccgcg	agaaccaggt	2280
ggagcgcgac	cgcgtgagca	tccccgccat	gcccacccag	atccagtacc	cccagtacca	2340
gcccgtggag	aacaagaccc	agccccctgg	ggtgtaccag	taccgcctgc	ccaccgagct	2400
gcagtaccgc	ccccccagcg	aggtcagta	ccgccccca	gcccgtgtgcc	ccgtgccccaa	2460
cagcaccgcc	cccttaccagc	agcccaccgc	catggccagc	aacagccccg	ccacccagga	2520
cgccgcccctg	tacccccagc	ccccccacgt	gcccctgaac	cccacccgcca	gccgcagcgg	2580
ccagggcggc	gccctgcacg	ccgtgatcga	cgaggcccgc	aagcagggcg	acctggaggc	2640
ctggcgcttc	ctgggtatcc	tgcaagctgg	gcaggccgc	gaggagaccc	agggtggcgc	2700
ccccgcccgc	gccgagaccc	gctgcgagcc	tttaccatg	aagatgtga	aggacatcaa	2760
ggagggcgtg	aagcagtacg	gcagcaacag	cccctacatc	cgcacccctgc	tggacagcat	2820
cgcccacggc	aaccgcctga	ccccctacga	ctgggagatc	ctggccaaga	gcagcctgag	2880
cagcagccag	tacctgcagt	tcaagacctg	gtggatcgac	ggcgtgcagg	agcaggtgcg	2940
caagaaccag	gccacccaagc	ccaccgtgaa	catgcacgcc	gaccagctgc	tgggcaccgg	3000
cccccaactgg	agcaccatca	accaggcagag	cgtgatgcag	aacaggggca	tcgagcaggt	3060
gchgccccatc	tgcctgcgcg	cctggggcaa	gatccaggac	cccggcaccg	ccttccccat	3120
caacagcatc	cgcaggggca	gcaaggagcc	ctacccgcac	ttcgtggccc	gcctgcagga	3180
cgcggcccg	aagagcatca	ccgacgacaa	cgcccccaag	gtatgcgtgg	agctgtatggc	3240
ctacgagaac	gccaaccccg	agtggcagag	cgccatcaag	cccctgaagg	gcaagggtgcc	3300
cgccggcgtg	gacgtatca	ccgagtagct	gaaggccctgc	gacggcatcg	gcggcggcat	3360
gcacaaggcc	atgctgtatgg	cccaggccat	gcgcggcctg	accctggggc	gccaggtgcg	3420
caccttcggc	aagaagtgt	acaactgcgg	ccagatcgcc	cacctgaagc	gcagctgccc	3480
cgtgctgaac	aagcagaaca	tcatcaacca	ggccatcacc	gccaagaaca	agaagcccag	3540
cggcctgtgc	cccaagtgcg	gcaagggcaa	gcactgggccc	aaccagtgcc	acagcaagtt	3600
cgacaaggac	ggccagcccc	tgagcggcaa	ccgcaagcgc	ggccagcccc	aggccccca	3660
gcagaccggc	gccttcccc	tgcaagctgtt	cgtgccccag	ggcttccagg	gccagcagcc	3720
cctgcagaag	atccccccccc	tgcaagggcgt	gagccagctg	cagcagagca	acagctgccc	3780
cgcggcccg	caggccgccc	cccaggctta	agaattcaga	ctcgagcaag	tctagaaaagc	3840
catggatatac	ggatccacta	cgcgttagag	ctcgctgatc	agcctcgact	gtgccttcta	3900
gttgccagcc	atctgttgtt	tgccccctccc	ccgtgccttc	cttgaccctg	gaagggtgcc	3960
ctccactgt	ccttccttaa	taaaatgagg	aaattgcattc	gcattgtctg	agtaggtgtc	4020
attctattct	gggggggtggg	gtggggcagg	acagcaaggg	ggaggattgg	gaagacaata	4080
gcaggggggt	gggcgaagaa	ctccagcatg	agatccccgc	gctggaggat	catccagccg	4140
gctgtcccgga	aaacgattcc	gaagcccaac	ctttcataga	aggcggcgggt	ggaatcgaaa	4200
tctcgatgt	gcaggttggg	cgtcgcttgg	tcggtcattt	cgaacccccag	agtcccgctc	4260
agaagaactc	gtcaagaagg	cgatagaagg	cgatgcgtg	cgaatcggtt	gcggcgtatac	4320
cgtaaagcac	gaggaagcgg	tcagccccatt	ccggccccaag	ctttcagca	atatcacggg	4380
tagccaaacgc	tatgtcctga	tagcggtccg	ccacacccag	ccggccacag	tcgatgaatc	4440
cagaaaaagcg	gccatTTTCC	accatgatata	tcggcaagca	ggcatcgcca	tgggtcacga	4500
cgagatccctc	gccgtcgccc	atgcgcct	tgagccttgc	gaacagtctg	gctggcgcga	4560
gccctgtatg	ctttcgttcc	agatcatcct	gatcgacaa	accggcttcc	atccgagatc	4620
gtgctcgctc	gatgcgtatgt	ttcgcttgg	ggtcgaatgg	gcaggttagcc	ggatcaagcg	4680
tatgcagccg	ccgcatttgc	tcagccatga	tggatactt	ctcggcaggaa	gcaagggttag	4740
atgacaggag	atccctcccc	ggcacttcgc	ccaaatagcg	ccagtcctt	cccgcttcag	4800
tgacaacgtc	gagcacagct	gwgcaaggaa	ccggccgtgt	ggccagccac	gatagcccg	4860
ctgcctcgctc	ctgcagttca	ttcagggcac	cgagacagtc	ggtttgaca	aaaagaaccg	4920
ggcgccccctg	cgctgacagc	cggaacacacgg	cggcattcaga	gcagccgatt	gtctgttgt	4980
cccagtccata	gccgaatagc	ctctccaccc	aagcggccgg	agaacctgcg	tgcaatccat	5040
tttgttcaat	catgcgaaac	gatccctcatc	ctgtcttttg	atcagatctt	gatccctgc	5100
gccatcagat	ccttggcggc	aagaaagcca	tccagtttac	tttgcaggggc	ttcccaacct	5160
taccagaggg	cgccccagct	ggcaattccg	gttcgttgc	tgtccataaa	accgcccagt	5220
ctagctatcg	ccatgttaagc	ccactgcaag	ctacctgtt	tctctttgcg	tttgcgtttt	5280
cccttgcgttca	gatagccccag	tagctgacat	tcatccgggg	tcagcaccgt	ttctgcggac	5340
tggctttctatc	cgtgttccgc	ttccctttagc	agcccttgcg	ccctgagtgc	ttgcggcagc	5400
gtgaagctaa	tatcatgttta	aatttttgg	aatatcgatc	attttttaaac	caataggccg	5460
aaatccggcaa	aatcccttat	aaatcaaaag	aatagcccg	gatagggttg	agtgttgttc	5520
cagtttggaa	caagagtc	ctattaaaga	acgtggactc	caacgtcaaa	ggcgaaaaaa	5580
ccgtctatca	gggcgtatggc	cggtatcagct	tatgcgggt	gaaataccgc	acagatgcgt	5640
aaggagaaaa	taccgcata	ggcgtcttcc	cgcttccctg	ctcactgact	cgctgcgttc	5700
ggtcgttccgg	ctgcggcggag	cggtatcagc	tcactcaa	gcccgtataac	ggttattccac	5760
agaatcagggg	gataacgcag	gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	5820
ccgtaaaaag	gccgcgttgc	ttggcgcccc	ccataggctc	cgccccccctg	acgagcatca	5880
caaaaatcg	cgctcaagtc	agaggtggcg	aaacccgaca	ggactataaa	gataccaggc	5940

Substitute Sequence Listing_USSN 10587032_PP019482.007

gtttccccct	ggaagctccc	tcgtgcgctc	tcctgttccg	accctgccgc	ttaccggata	6000
cctgtccgcc	tttctccctt	cgggaagcgt	ggcgcttct	catagctcac	gctgttagta	6060
tctcagttcg	gtgttaggtcg	ttcgctccaa	gctgggctgt	gtgcacgaac	cccccggtca	6120
gccccaccgc	tgcgccctat	ccggtaacta	tcgtctttag	tccaaacccgg	taagacacga	6180
cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcggaggt	atgtaggcgg	6240
tgctacagag	ttcttgaagt	ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	6300
tatctgcgt	ctgctgaagc	cagtacctt	cggaaaaaga	gttggtagct	cttgatccgg	6360
caaacaacc	accgctggta	gcgggtggtt	tttggttgc	aagcagcaga	ttacgcgcag	6420
aaaaaaaaagg	tctcaagaag	atccctttagt	cttttctact	gaacgggtat	ccccaccgg	6480
attgcg						6486

<210> 81
<211> 2103
<212> DNA

<213> Human endogenous retrovirus, K family (HERV-K)

<400>	81					
atgaacccaa	gcgagatgca	aagaaaagca	cctccgcgga	gacggagaca	tcgcaatcga	60
gcaccgttga	ctcacaagat	gaacaaaatg	gtgacgtcag	aagaacagat	gaagttgcca	120
tccaccaaga	aggcagagcc	gccaaactgg	gcacaactaa	agaagctgac	gcagttagct	180
acaaaatatac	tagagaacac	aaagtgcaca	caaaccgg	agatgtatgc	gcttgcagcc	240
ttgatgattt	tatcaatgtt	ggttaagtctc	cctatgcctg	caggagcagc	tgcagcta	300
tataacctact	gggcctatgt	gccttcccg	cccttaattt	gggcgtcac	atggatggat	360
aatccctacag	aagtatatgt	taatgatagt	gtatgggtac	ctggcccccac	agatgatcgc	420
tgccctgcca	aacctgagga	agaagggatg	atgataaata	tttccatttgg	gtatcattat	480
cctccatttt	gccttagggag	agcaccagga	tgtttaatgc	ctgcagtc	aaattgttg	540
gtagaagtac	ctactgtcag	tcccatctgt	agattcactt	atcacatgg	aagcggatg	600
tcactcaggc	cacggtaaaa	ttatttacaa	gacttttctt	atcaaagatc	attaaaattt	660
agacctaaag	ggaaacctt	ccccaaaggaa	attcccaag	aatcaaaaaaa	tacagaagtt	720
ttagtttggg	aagaatgtgt	ggccaaatgt	gcggtgat	tacaaaacaa	tgaattcga	780
actattatag	attgggcacc	tcgaggtcaa	ttctaccaca	attgctcagg	acaaactcag	840
tcgtgtccaa	gtgcacaagt	gagttccagct	gttgatagcg	acttaacaga	aagtttagac	900
aaacataagc	ataaaaaaatt	gcagtcttc	tacccttggg	aatggggaga	aaaaggaatc	960
tctaccccaa	gaccaaaaat	agtaagtcc	gtttctgtc	ctgaacatcc	agaattatgg	1020
aggcttactg	tggcttcaca	ccacattaga	atttggtctg	gaatcaaac	tttagaaaca	1080
agagatcgta	agccattttt	tactattgac	ctgaatttca	gtctaacagt	tcctttacaa	1140
agttgcgtaa	agccccctt	tatgttagtt	gttagaaata	tagttattaa	accagactcc	1200
cagactataa	cctgtgaaaa	ttgttaggt	cttacttgc	ttgattcaac	tttaatttgg	1260
caacacgtt	tttgcgttgt	gagagcaaga	gagggcggt	ggatccctgt	gtccatggac	1320
cgaccgtggg	aggcctcgcc	atccgttcat	atttgactg	aagtattaaa	aggtgtttt	1380
aatagatcca	aaagattcat	ttttacttta	attgcagtg	ttatgggtt	aattgcagtc	1440
acagctacgg	ctgctgttagc	aggagttgc	ttgcactt	ctgttcagtc	agtaaaactt	1500
gttaatgatt	ggcaaaaaaaa	ttctacaaga	ttgtggatt	cacaatctag	tattgatcaa	1560
aaattggcaa	atcaaattaa	tgatcttga	caaactgtca	tttggatggg	agacagactc	1620
atgagcttag	aacatcgtt	ccagttacaa	tgtgactgga	atacgtcaga	ttttgtatt	1680
acaccccaa	tttataatga	gtctgagcat	cactggac	ttgttagacg	ccatctacag	1740
ggaagagaag	ataatctcac	tttagacatt	tccaaattaa	aagaacaaat	tttcgaagca	1800
tcaaaagccc	attnaaattt	ggtgccagga	actgaggca	ttgcaggagt	tgctgtggc	1860
ctcgcaaatc	ttaaccctgt	cacttgggtt	aagaccattt	gaagtactac	gattataaat	1920
ctcatattaa	tccttgcgtg	cctttttgt	ctgtgttag	tctgcagggt	tacccaaacag	1980
ctccgaagag	acagcgacca	tcgagaacgg	gccatgtat	cgatggcggt	tttgcgaaa	2040
agaaaagggg	gaaatgtggg	gaaaagcaag	agagatcaga	ttgttactgt	gtctgtggcc	2100
taa						2103

<210> 82
<211> 2103
<212> DNA
<213> Artificial Sequence

<220>
<223> Modified env sequence

<400> 82

Substitute Sequence Listing_USSN 10587032_PP019482.007

atgaacccca	gcgagatgca	gcfcaaggcc	cccccccgcc	gcccggcca	ccgcaaccgc	60
gccccctga	cccacaagat	gaacaagatg	gtgaccagcg	aggagcagat	gaagctgcc	120
agcaccaaga	aggccgagcc	ccccacctgg	gcccagctga	agaagctgac	ccagctggcc	180
accaagtacc	tggagaacac	caagggtgacc	cagaccccg	agagcatgct	gctggccgccc	240
ctgatgatcg	tgagcatggt	ggtggacctg	cccatgccc	ccggcggcc	cgccgccaac	300
tacacctact	gggcctacgt	gcccattcccc	cccccgtatcc	gcccgtgac	ctggatggac	360
aaccccaaccg	aggtgtacgt	gaacgacagc	gtgtgggtgc	ccggccccat	cgacgaccgc	420
tgccccgcca	agcccgagga	ggagggcatg	atgatcaaca	tcagcatcg	ctaccactac	480
ccccccatct	gcctgggccc	cgccccccggc	tgcctgtatgc	ccggcgtgca	gaactggctg	540
gtggaggtgc	ccaccgttag	ccccatctgc	cgcttcaccc	accacatggt	gagccgcatg	600
agcctgcgcc	cccgctgaa	ctacctgcag	gacttcagct	accagcgcag	cctgaagttc	660
cgccccaagg	gcaagccctg	ccccaaaggag	atccccaaagg	agagcaagaa	caccgaggtg	720
ctgggtgtggg	aggagtgcgt	ggccaaacagc	gccgtatcc	tgcagaacaa	cgagttcggc	780
accatcatcg	actggggccc	ccgcggccag	ttctaccaca	actgcagcgg	ccagaccag	840
agctgccc	gcccggcagg	gagccccggc	gtggacagcg	acctgaccga	gagcctggac	900
aagcacaa	acaagaagct	gcagagcttc	tacccctggg	agtggggcga	gaagggcattc	960
agcacccccc	gcccccaagat	cgtgagcccc	gtgagcggcc	ccgagcaccc	cgagctgtgg	1020
cgcctgaccg	tggccagcc	ccacatccgc	atctggagcg	gcaaccagac	cctggagacc	1080
cgcgcggca	agcccttcta	caccatgcac	ctgaacagca	gcctgaccgt	gccccctgcag	1140
agctgcgtg	agcccccccta	catgtgggt	gtggcaaca	tcgtatcaa	gcccgcacgc	1200
cagaccatca	cctgcgagaa	ctgcccctgc	ctgacactgc	tcgacagcac	cttcaactgg	1260
cagaccgc	tcctgctgg	gcccggccgc	gagggcgtgt	ggatccccgt	gagcatggac	1320
cgccttggg	aggccagccc	cagcgtgcac	atcctgaccg	aggtgctgaa	gggcgtgctg	1380
aaccgcagca	agcgcttcat	ttcacccctg	atcgccgtga	tcatgggcct	gatcgccgtg	1440
accgcaccg	ccgcccgtgg	ccgcgtggcc	ctgcacagca	gcgtgcagag	cgtgaacttc	1500
gtgaacgact	ggcagaagaa	cagcaccgc	ctgtggaaaca	gccagagcag	catcgaccag	1560
aagctggcca	accagatcaa	cgaccgtgc	cagaccgtga	tctggatggg	cgaccgcctg	1620
atgagcctgg	agcaccgctt	ccagctgcag	tgcgacttgg	acaccagcga	cttctgcattc	1680
accccccaga	tctacaacga	gagcgagcac	cactggaca	tggtgccgc	ccacctgcag	1740
ggccgcgagg	acaacctgac	cctggacatc	agcaagctga	aggagcagat	cttcgaggcc	1800
agcaaggccc	acctgaacct	ggtgcccggc	accgaggcca	tcgcccggcgt	ggccgacggc	1860
ctggccaacc	tgaacccctgt	gaccctgggt	aagaccatcg	gcagcaccac	catcatcaac	1920
ctgatcctga	tcctgggtgt	cctgttctgc	ctgctgtgg	tgtggccctg	caccgcag	1980
ctgcggccgc	acagcgacca	ccgcgagcgc	gccatgtga	ccatggccgt	gctgagcaag	2040
cgcaaggcg	gcaacgtggg	caagagcaag	cgcgaccaga	tcgtgaccgt	gagcgtggcc	2100
taa						2103

<210> 83

<211> 700

<212> PRT

<213> Human endogenous retrovirus, K family (HERV-K)

<400> 83

Met	Asn	Pro	Ser	Glu	Met	Gln	Arg	Lys	Ala	Pro	Pro	Arg	Arg	Arg	Arg
1					5			10				15			

His	Arg	Asn	Arg	Ala	Pro	Leu	Thr	His	Lys	Met	Asn	Lys	Met	Val	Thr
					20				25			30			

Ser	Glu	Glu	Gln	Met	Lys	Leu	Pro	Ser	Thr	Lys	Lys	Ala	Glu	Pro	Pro
				35		40				45					

Thr	Trp	Ala	Gln	Leu	Lys	Lys	Leu	Thr	Gln	Leu	Ala	Thr	Lys	Tyr	Leu
					50		55			60					

Glu	Asn	Thr	Lys	Val	Thr	Gln	Thr	Pro	Glu	Ser	Met	Leu	Leu	Ala	Ala
					65		70		75		80				

Leu	Met	Ile	Val	Ser	Met	Val	Val	Ser	Leu	Pro	Met	Pro	Ala	Gly	Ala
					85		90				95				

Ala	Ala	Ala	Asn	Tyr	Thr	Tyr	Trp	Ala	Tyr	Val	Pro	Phe	Pro	Pro	Leu
					100		105				110				

Substitute Sequence Listing_USSN 10587032_PP019482.007

Ile Arg Ala Val Thr Trp Met Asp Asn Pro Thr Glu Val Tyr Val Asn
115 120 125

Asp Ser Val Trp Val Pro Gly Pro Ile Asp Asp Arg Cys Pro Ala Lys
130 135 140

Pro Glu Glu Glu Gly Met Met Ile Asn Ile Ser Ile Gly Tyr His Tyr
145 150 155 160

Pro Pro Ile Cys Leu Gly Arg Ala Pro Gly Cys Leu Met Pro Ala Val
165 170 175

Gln Asn Trp Leu Val Glu Val Pro Thr Val Ser Pro Ile Cys Arg Phe
180 185 190

Thr Tyr His Met Val Ser Gly Met Ser Leu Arg Pro Arg Val Asn Tyr
195 200 205

Leu Gln Asp Phe Ser Tyr Gln Arg Ser Leu Lys Phe Arg Pro Lys Gly
210 215 220

Lys Pro Cys Pro Lys Glu Ile Pro Lys Glu Ser Lys Asn Thr Glu Val
225 230 235 240

Leu Val Trp Glu Glu Cys Val Ala Asn Ser Ala Val Ile Leu Gln Asn
245 250 255

Asn Glu Phe Gly Thr Ile Ile Asp Trp Ala Pro Arg Gly Gln Phe Tyr
260 265 270

His Asn Cys Ser Gly Gln Thr Gln Ser Cys Pro Ser Ala Gln Val Ser
275 280 285

Pro Ala Val Asp Ser Asp Leu Thr Glu Ser Leu Asp Lys His Lys His
290 295 300

Lys Lys Leu Gln Ser Phe Tyr Pro Trp Glu Trp Gly Glu Lys Gly Ile
305 310 315 320

Ser Thr Pro Arg Pro Lys Ile Val Ser Pro Val Ser Gly Pro Glu His
325 330 335

Pro Glu Leu Trp Arg Leu Thr Val Ala Ser His His Ile Arg Ile Trp
340 345 350

Ser Gly Asn Gln Thr Leu Glu Thr Arg Asp Arg Lys Pro Phe Tyr Thr
355 360 365

Ile Asp Leu Asn Ser Ser Leu Thr Val Pro Leu Gln Ser Cys Val Lys
370 375 380

Pro Pro Tyr Met Leu Val Val Gly Asn Ile Val Ile Lys Pro Asp Ser
385 390 395 400

Gln Thr Ile Thr Cys Glu Asn Cys Arg Leu Leu Thr Cys Ile Asp Ser
405 410 415

Thr Phe Asn Trp Gln His Arg Ile Leu Leu Val Arg Ala Arg Glu Gly
420 425 430

Val Trp Ile Pro Val Ser Met Asp Arg Pro Trp Glu Ala Ser Pro Ser
435 440 445

Substitute Sequence Listing_USSN 10587032_PP019482.007

Val His Ile Leu Thr Glu Val Leu Lys Gly Val Leu Asn Arg Ser Lys
450 455 460

Arg Phe Ile Phe Thr Leu Ile Ala Val Ile Met Gly Leu Ile Ala Val
465 470 475 480

Thr Ala Thr Ala Ala Val Ala Gly Val Ala Leu His Ser Ser Val Gln
485 490 495

Ser Val Asn Phe Val Asn Asp Trp Gln Lys Asn Ser Thr Arg Leu Trp
500 505 510

Asn Ser Gln Ser Ser Ile Asp Gln Lys Leu Ala Asn Gln Ile Asn Asp
515 520 525

Leu Arg Gln Thr Val Ile Trp Met Gly Asp Arg Leu Met Ser Leu Glu
530 535 540

His Arg Phe Gln Leu Gln Cys Asp Trp Asn Thr Ser Asp Phe Cys Ile
545 550 555 560

Thr Pro Gln Ile Tyr Asn Glu Ser Glu His His Trp Asp Met Val Arg
565 570 575

Arg His Leu Gln Gly Arg Glu Asp Asn Leu Thr Leu Asp Ile Ser Lys
580 585 590

Leu Lys Glu Gln Ile Phe Glu Ala Ser Lys Ala His Leu Asn Leu Val
595 600 605

Pro Gly Thr Glu Ala Ile Ala Gly Val Ala Asp Gly Leu Ala Asn Leu
610 615 620

Asn Pro Val Thr Trp Val Lys Thr Ile Gly Ser Thr Thr Ile Ile Asn
625 630 635 640

Leu Ile Leu Ile Leu Val Cys Leu Phe Cys Leu Leu Leu Val Cys Arg
645 650 655

Cys Thr Gln Gln Leu Arg Arg Asp Ser Asp His Arg Glu Arg Ala Met
660 665 670

Met Thr Met Ala Val Leu Ser Lys Arg Lys Gly Gly Asn Val Gly Lys
675 680 685

Ser Lys Arg Asp Gln Ile Val Thr Val Ser Val Ala
690 695 700